RICE UNIVERSITY

Structural and Functional Studies on the Infectious Salmon Anemia
Virus Nucleoprotein

By

Wenjie Zheng

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Yizhi Jane Tao, Ph.D., Associate Professor, Advisor
Biochemistry and Cell Biology

Kathleen Beckingham, Ph.D., Professor
Biochemistry and Cell Biology

Edward Nikonowicz, Ph.D., Associate Professor
Biochemistry and Cell Biology

Junghae Suh, Ph.D., Assistant Professor
Bioengineering

Qinghua Wang, Ph.D., Assistant Professor
Biochemistry and Molecular Biology
Baylor College of Medicine

HOUSTON, TEXAS
MAY 2013
Abstract

Structural and Functional Studies on the Infectious Salmon Anemia Virus Nucleoprotein

By

Wenjie Zheng

Genome packaging for viruses with segmented genomes is often a complex problem. This is particularly true for influenza viruses and other orthomyxoviruses which are able to cause infectious disease, and even worldwide pandemics. The genome of Orthomyxovirus consists of 6-8 negative-sense RNAs encapsidated as ribonucleoprotein (RNP) complexes which perform multiple essential functions throughout the virus life cycle. To better understand the structural features of orthomyxovirus RNPs that allow them to be specifically packaged, we performed structural/functional studies of the nucleoprotein (NP), the major protein component of the RNP, from the infectious salmon anemia virus (ISAV). The crystal structure of the ISAV-NP was determined to 2.7Å resolution. The ISAV-NP possesses a 112-aa (amino acid) N-terminal domain and a bi-lobular core structure that strongly resembles the structure of the influenza virus NP. Because the ISAV-NP forms homogenous dimers that are stable in solution, I was able to study the NP:RNA binding affinity as well as stoichiometry with fluorescence polarization, using recombinant proteins and synthetic oligos. Surprisingly, the RNA binding analysis revealed that each NP binds ~12 nts (Nucleotide) of RNA, shorter than the 24-28 nts...
originally estimated for the influenza A virus NP. The 12-nt stoichiometry was further confirmed by results from electron microscopy and dynamic light scattering. These results suggest that free RNA exists in the orthomyxovirus RNPs, and selective RNP packaging is likely accomplished through direct RNA-RNA interactions.
Acknowledgement

The work presented herein would not have been possible without the help, support and understanding from many people, including my adviser, professors, colleagues, friends, funding agencies, and my families. They have made my graduate studies one of the most rewarding and memorable experiences in my life. For them, I am very grateful.

Dr. Yizhi Tao has been the best mentor I have ever had. Her passion for science, demand for excellence, and insights on the big-pictures have deeply impressed me. She always serves as an excellent model of scientist, by demonstrating solid grasp of knowledge, consistent diligence for success, great attention to details, and eternal pursuit of the truth. I could never have gone this far without her patient guidance and thought-provoking directions, especially during the many frustrations and setbacks during the past years. She is a great mentor, who values critical thinking and thorough consideration much more than simply arriving at the correct answers. Instead of issuing “what-to-do” orders, she always leads me to develop the ability on “how-to-think”. She is a talented team leader with superior professionalism and managerial skills, who has created a harmonious and cooperative lab atmosphere that benefits and encourages every lab member. She is even a considerate friend, who is always willing to listen, help, and keep me optimistic, nurturing this project from the infancy. Thanks to Dr. Tao’s guidance and help, I was able to improve myself on many aspects: science, logic, professionalism, and even English-writing. I truly appreciate all her help and support.
I would like to express my enormous gratefulness to my progress review committee professors: Dr. Edward Nikonowicz, Dr. Kathleen Beckingham, and Dr. Qinghua Wang. They deserve the top praise for creating a flexible environment of my research, providing insights and feedback on scientific issues, and guiding me through my evolution as a scientist. Dr. Nikonowicz has been on my progress review committee since I started. His valuable suggestions and insights for my project, especially on the biophysical characterization of the protein-RNA interactions, have been proven very helpful. Dr. Beckingham has also been very generous on helping me to push my project forward. In particular, she suggested strategies to design different control experiments to examine the mutated proteins, which have made all my experiments be a complete story. Dr. Wang, as an expert on the influenza viruses, has guided me to approach the same virology issue from different angles, which greatly broadened my views in understanding my project. Also, I thank Dr. Junghae Suh for agreeing to serve as the outside member on my thesis committee. Finally, I would also thank Dr. Michael Gustin and Dr. Kevin Mackenzie, who were on my progress review committee in the early years and contributed to my research.

I would like to thank Dr. John Olson for his generous help and support on many aspects during the past years. Whenever I had questions about quantitative analysis of protein-ligand interaction, Dr. Olson was there to help. Dr. Olson contributed significantly in my research, by providing many useful insights and discussions. I am also grateful for our collaborator Dr. Vikram Vakharia who provided us the genetic materials
for this project.

I am also grateful for all the current and previous Tao lab members: Dr. Junhua Pan, Dr. Yunyun Jiang, Mr. Aaron Collier, Dr. Yukimatsu Toh, Ms. Yangyang Dong, Dr. Li Lin, Dr. Tom Guu, Mr. Haijiang Chen, Ms. Yusong Guo, Mr. Justin Harper, Dr. Qiaozhen Ye, Dr. Jinhui Dong, and Ms. Liping Dong. I am very thankful for Dr. Junhua Pan who taught me protein purification skills, crystallization methods, and other biophysical techniques. He also contributed a lot in guiding me to develop a mindset to think as a scientist. I also would like to thank Dr. Yunyun Jiang for sharing lab techniques, being an amazing friend, as well as offering generous help when I first came to the United States. I thank Mr. Aaron Collier for helping countless times on my English-writing, aiding me with the CD and the Mass Spec experiments, and most importantly, being such a wonderful friend. I appreciate Dr. Yukimatsu Toh’s help with my research, including useful discussions and hands-on guidance, especially during the structure determination of my protein.

I thank the Shamoo Lab, the Nikonowicz lab, the Olson lab, and the Silberg lab for sharing reagents and equipments. I am thanks to Dr. Wenhua Guo for offering training sessions and technical supports with regard to the electron microscopy. I am grateful to Mr. Richard Crouse for maintaining the X-ray diffraction facility and solving technical problems. I appreciate that Dr. Jinhui Dong taught me the skills on fluorescence anisotropy. I thank Yangyang Dong’s help with molecular cloning. I am thankful for the
staff at Cornell High Energy Synchrotron Source (CHESS), Advanced Photon Source (APS), and Advanced Light Source (ALS), for helping me with data collection.

I thank the Department of Biochemistry and Cell Biology, especially Dr. Janet Braam, Ms. Sandra Saunders, Dr Susan Cates, Ms. Angel Forward, Ms. Monica Roberts, Mr. Juan Sanchez, and Ms. Minnie Taylor, who are always willing to help me whenever I have any kind of issues. I cherish the joyful and memorable moments with the fellow graduate students from the department, especially Mr. Aaron Collier, Dr. Matt Pena, Dr. Zhao Huang, Mr. John Liu, Dr. Andres Benitez, Mr. Corwin Miller, Dr. Avani Verma, Mr. Joseph Faust, and Mr. Andrew Chang. I thank the help from the staff at the Gulf Coast Consortia (GCC) who provided help during my fellowship of Nanobiology Interdisciplinary Graduate Training Program.

In the end, I would like to express my deep gratitude to my wife, Yuanyuan Li, my parents, and my grandparents. I sincerely appreciate Yuanyuan for her unyielding love, unconditional support, and great patience during the past years. I could never go this far without her comfort and encouragement. Also, I would like to thank my parents and my grandparents, for always believing in me and supporting my pursuit, and for their great sacrifices to keep me focused on my research. To my families, I owe too many thanks.

I am truly grateful for everyone who has ever helped me during my graduate studies at Rice. I sincerely apologize if I have unintentionally left anyone’s name out.
INDEX

1. Introduction ................................................................. 1
   1.1 Orthomyxovirus ......................................................... 1
   1.2 Influenza A virus ....................................................... 2
      1.2.1 Overview ............................................................. 2
      1.2.2 Viral Life cycle ................................................... 5
   1.3 The ribonucleoprotein (RNP) complex ............................ 7
      1.3.1 The components of the RNP .................................... 7
      1.3.2 The RNP Structure ................................................ 18
      1.3.3 RNP in viral transcription and replication .................. 22
      1.3.4 The RNP assembly ................................................ 25
      1.3.5 The RNP trafficking .............................................. 28
      1.3.6 Specific RNP packaging ......................................... 30
   1.4 Infectious Salmon Anemia Virus .................................... 34
      1.4.1 Overview ............................................................. 34
      1.4.2 ISAV proteins ...................................................... 38
      1.4.3 ISAV can be a good model for studies of the influenza A virus .... 41
   1.5 Anti-flu drugs targeting influenza proteins ....................... 42

2. Methods ........................................................................... 45
   2.1 Plasmid construction .................................................... 47
   2.2 Protein expression ....................................................... 49
   2.3 Protein purification ..................................................... 51
   2.4 Crystallization ............................................................ 55
   2.5 Crystal manipulation .................................................... 59
      2.5.1 Crystal annealing .................................................. 59
      2.5.2 Crystal dehydration ............................................... 60
      2.5.3 Crystal cross-linking .............................................. 61
      2.5.4 Seeding ............................................................... 62
   2.6 X-ray crystallography ................................................... 64
      2.6.1 Data collection ..................................................... 64
      2.6.2 Phasing ............................................................... 71
      2.6.3 Structure determination ......................................... 75
   2.7 Biophysical experiments ................................................. 81
      2.7.1 Fluorescence polarization ....................................... 81
      2.7.2 Electron Microscopy ............................................. 84
      2.7.3 Dynamic Light Scattering ..................................... 86

3. Production and Characterization of the ISAV-NP ..................... 88
   3.1 Cloning ...................................................................... 88
   3.2 Protein expression ....................................................... 89
   3.3 Protein purification ..................................................... 91
   3.4 Biochemical Characterization of the ISAV-NP .................... 93
      3.4.1 Degradation test .................................................. 94
      3.4.2 Oligomerization test ............................................. 94
3.4.3 RNA-binding test ........................................................................................................... 95
3.5 Crystallization of ISAV-NP ............................................................................................... 96
3.6 Lysine methylation of ISAV-NP ....................................................................................... 98
4. Production and Characterization of the ISAV-NP truncation mutants ...................... 100
  4.1 Construct design .............................................................................................................. 100
  4.1.1 Limited proteolysis .................................................................................................... 100
  4.1.2 Primary sequence analysis ....................................................................................... 103
  4.1.3 Secondary structure analysis .................................................................................. 103
  4.2 Expression, purification and crystallization of the ISAV-NP truncations ......... 105
  4.3 Summary of the ISAV-NP truncations ........................................................................... 112
5. The assembly of the ISAV-NP and nucleic acid complexes ............................................. 113
  5.1 Strategy for forming the complex ................................................................................. 113
      5.1.1 Choice of protein .................................................................................................. 114
      5.1.2 Choice of RNA ..................................................................................................... 116
      5.1.3 Other parameters ................................................................................................. 118
  5.2 Co-crystallization of ISAV-NPΔC16 and nucleic acids ................................................. 119
6. Data collection and structure determination ................................................................. 121
  6.1 Improve ISAV-NP crystal by crystal manipulation ...................................................... 121
  6.2 Data collection ............................................................................................................... 122
  6.3 Structure determination ............................................................................................... 123
7. The crystal structure of the ISAV-NP ............................................................................... 133
  7.1 Dimerization via the tail loop ....................................................................................... 133
  7.2 The RNA-binding groove ......................................................................................... 137
  7.3 The N-terminal domain ............................................................................................. 141
  7.4 The missing density of nucleic acids ....................................................................... 143
  7.5 ISAV-NP is a good model to study the RNA-binding ................................................. 145
8. Characterization of the RNA binding of the ISAV-NP ............................................... 146
  8.1 Fluorescence Polarization ......................................................................................... 146
  8.2 Electron Microscopy ................................................................................................... 154
  8.3 Dynamic Light Scattering ........................................................................................... 156
  8.4 Functional implication ............................................................................................... 157
9. Major conclusions ............................................................................................................ 159
10. Future directions ............................................................................................................ 162
    10.1 Structure determination of the ISAV-NP-RNA complex ......................................... 162
    10.2 Structure determination of the ISAV-NP monomer ............................................... 164
Supplementary information .......................................................................................... 166
S1. Structural studies on other ISAV proteins ................................................................. 166
  S1.1 ISAV-M1 .................................................................................................................. 166
  S1.2 ISAV-M2 ................................................................................................................ 171
  S1.3 ISAV-NSI ............................................................................................................... 173
  S1.4 ISAV-NEP ............................................................................................................. 174
S2 Program for data analysis ............................................................................................ 175
S3 Genes and primer sequences ....................................................................................... 177
Bibliography ..................................................................................................................... 194
Illustrations

1. Introduction
   Figure 1.1 Structure of the influenza A virus .......................................................... 2
   Figure 1.2 Cartoon model of the RNP organization ...................................................... 4
   Figure 1.3 Life cycle of the influenza A virus ............................................................. 5
   Figure 1.4 The NP structures of the influenza viruses ............................................... 10
   Figure 1.5 The crystal structures of the influenza A virus polymerase complex .......... 14
   Figure 1.6 The influenza A virus RNP structure ......................................................... 19
   Figure 1.7 RNP replication and transcription of the influenza A virus ...................... 24
   Figure 1.8 Specific packaging of the influenza A virus RNP ..................................... 31
   Figure 1.9 Electron microscopic graphs of the influenza A virus and the ISAV ....... 34
   Figure 1.10 Electron microscopic images of the ISAV RNPS ...................................... 35
   Figure 1.11 Schematic drawing of the ISAV life cycle ............................................... 37
   Figure 1.12 Structures of sialic acid, zanamivir, and oseltamivir ............................. 42

2. Methods
   Figure 2.1 Procedures to determine a protein structure from its gene ...................... 45
   Figure 2.2 FPLC instrument and columns for chromatography .................................. 52
   Figure 2.3 Protein crystallization .............................................................................. 56
   Figure 2.4 Crystallization instruments and tools ......................................................... 57
   Figure 2.5 Crystal dehydration ................................................................................... 60
   Figure 2.6 Crystal cross-linking ................................................................................. 61
   Figure 2.7 X-ray beams reflection .............................................................................. 64
   Figure 2.8 Ewald's sphere ......................................................................................... 65
   Figure 2.9 Bravais lattices .......................................................................................... 67
   Figure 2.10 Fluorescence polarization .......................................................................... 81
   Figure 2.11 Negative-staining electron microscopy ..................................................... 84

3. Production and Characterization of ISAV-NP
   Figure 3.1 Cloning of ISAV-NP ................................................................................. 89
   Figure 3.2 Expression tests of ISAV-NP at different temperatures .......................... 90
   Figure 3.3 ISAV-NP purification with Ni-NTA resins ................................................ 91
   Figure 3.4 ISAV-NP purification with Heparin column ............................................. 92
   Figure 3.5 ISAV-NP purification with size-exclusion chromatography .................. 93
   Figure 3.6 Degradation test of purified ISAV-NP ....................................................... 94
   Figure 3.7 Oligomerization test of purified ISAV-NP ................................................. 95
   Figure 3.8 Gel-filtration analysis of ISAV-NP with and without RNA addition ....... 95
   Figure 3.9 Crystallization of ISAV-NP ................................................................. 96
   Figure 3.10 Lysine methylation of ISAV-NP .............................................................. 98
4. Production and Characterization of ISAV-NP mutants

- Figure 4.1 Limited proteolysis of ISAV-NP ................................................................. 101
- Figure 4.2 Time-series digestion profile of ISAV-NP by trypsin ................................. 101
- Figure 4.3 Identification of the ISAV-NP truncation ΔN111 ........................................ 102
- Figure 4.4 Secondary structure prediction of ISAV-NP ............................................... 104
- Figure 4.5 Expression of ISAV-NP ΔN189 .................................................................. 105
- Figure 4.6 Purification, crystallization and characterization of ISAV-NP ΔN111 ......... 107
- Figure 4.7 Expression tests of ISAV-NP ΔN26 ........................................................... 108
- Figure 4.8 Purification, crystallization and characterization of ISAV-NP ΔC16 ............ 109
- Figure 4.9 Size-exclusion chromatography of N-terminal truncations ......................... 110
- Figure 4.10 Crystals of ISAV NP ΔN50, ΔN64, and ΔN84 ............................................ 111

5. The assembly of the ISAV-NP and nucleic acid complexes

- Figure 5.1 Size-exclusion chromatographic curves of ISAV-NP:DNA complexes .......... 117
- Figure 5.2 Diffraction of ISAV-NP ΔC16 crystals ....................................................... 120

6. Data collection and structure determination

- Figure 6.1 The Bravais lattice table of ISAV-NP crystal generated with HKL2000 ......... 124
- Figure 6.2 The reflection image of ISAV-NP crystal after refinement ......................... 125
- Figure 6.3 The final refinement statistics of ISAV-NP crystal from HKL2000 .......... 125
- Figure 6.4 The user-interface of Autosol and the parameters supplied ....................... 128
- Figure 6.5 The secondary structure prediction of the N terminal domain ................. 129
- Figure 6.6 The electron density map of the N-terminal β-turn-β-turn-β structure ..... 129
- Figure 6.7 The electron density of Arg 20 and Arg 21 with residues built in ............... 130
- Figure 6.8 The B-factor plot generated by Phenix ...................................................... 131

7. The crystal structure of ISAV-NP

- Figure 7.1 ISAV-NP crystal structure ......................................................................... 134
- Figure 7.2 ISAV-NP monomer with secondary structures labeled ............................ 135
- Figure 7.3 Structure comparison of ISAV-NP and influenza A virus NP .................. 136
- Figure 7.4 Size-exclusion chromatograms of ISAV-NP proteins ............................... 138
- Figure 7.5 Calculated electronic potential of an NP monomer .................................. 139
- Figure 7.6 ISAV-NP secondary structure assignment ............................................... 140

8. Characterization of the RNA binding of ISAV-NP

- Figure 8.1 The binding affinity measurements between RNAs and ISAV-NP ............ 148
- Figure 8.2 RNA binding stoichiometry measurements for the wt NP ...................... 151
- Figure 8.3 RNA binding stoichiometry measurements for the monomeric NP .......... 152
- Figure 8.4 Circular dichroism characterization of different ISAV NP proteins ........... 153
- Figure 8.5 ISAV-NP:RNA complexes characterized by electron microscopy ........... 155
- Figure 8.6 Molecular weights of NP and NP:RNA complexes measured by DLS ..... 156
S1. Structural studies on other ISAV proteins

Figure S1.1  Purification of ISAV-M1 with size-exclusion chromatography.......... 166
Figure S1.2  Crystals of ISAV-M1...................................................................... 167
Figure S1.3  X-ray diffraction of SeMet-M1 crystals ........................................ 168
Figure S1.4  Purification of ISAV-M2 with Size-exclusion chromatography......... 171
Figure S1.5  Purification of ISAV-NS1 with Size-exclusion chromatography........ 173
Figure S1.6  Expression test of ISAV-NEP............................................................. 174

S2. Program for data analysis

Figure S2.1  The graphical user interface of the Data Analysis tool ..................... 175
Figure S2.2  Read in and display data................................................................. 176
Figure S2.3  Modelling, plotting, and statistical testing ........................................ 176
Tables

4. Production and Characterization of ISAV-NP mutants
   Table 4.1 Summary of ISAV-NP full length protein and all ISAV-NP truncations.....112

5. The assembly of the ISAV-NP and nucleic acid complexes
   Table 5.1 Summary of ISAV-NP full length protein and all ISAV-NP truncations.....114
   Table 5.2 Comparison of the RNA-binding affinity of wt ISAV NP and ΔC16 .....115

6. Data collection and structure determination
   Table 6.1 The $R_{merge}$ and I/σ values for each resolution bin .......................... 126
   Table 6.2 The completeness of each resolution bin ............................................. 127
   Table 6.3 The redundancies of each resolution bin.............................................. 127
   Table 6.4 Statistics of Data collection, phasing, and model building .................... 132

S1. Structural studies on other ISAV proteins
   Table S1.1 Statistics of data collection of ISAV SeMet-M1 crystal ......................... 168
   Table S1.2 Peaks from Patterson map calculated from the M1 data ...................... 169
Acronyms

ISAV: Infectious Salmon Anemia Virus
RNP: ribonucleoprotein
NP: nucleoprotein
HA: hemagglutinin
HE: hemagglutinin -esterase
NA: neuraminidase
M1: matrix protein 1
M2: matrix protein 2
RdRp: RNA-dependent RNA-polymerase
PB1: polymerase basic protein 1
PB2: polymerase basic protein 2
PA: polymerase acidic protein
NEP: nuclear export protein
NS1: non-structural proteins 1
NS2: non-structural proteins 2
aa: amino acid
nt: nucleotide
FPLC: Fast Protein Liquid Chromatography
IPTG: Isopropyl β-D-1-thiogalactopyranoside
mRNA: messenger RNA
vRNA: virion RNA
vRNP: viral ribonucleoprotein
cRNA: complementary RNA
ssRNA: single-stranded RNA
dsRNA: double-stranded RNA
MAD: Multi-wavelength anomalous dispersion
SAD: Single-wavelength anomalous dispersion
MIR: Multiple isomorphous replacement
MR: Molecular replacement
SeMet: selenomethionine
CD: Circular dichroism
FA: Fluorescence Anisotropy
FP: Fluorescence polarization (same as FA)
DLS: Dynamic light scattering
EM: Electron microscopy
1. Introduction

1.1 Orthomyxovirus

The Orthomyxoviruses are a family of viruses whose genomes consist of segmented, single-stranded, and negative-sense RNAs. There are five different genera in this virus family: the influenza viruses A, B, and C, Thogotovirus, and Isavirus (Palese and Shaw, 2007). In 2009, a potential sixth genus of the Orthomyxoviridae family, made up of the Quaranfil, Johnston Atoll, and Lake Chad Viruses, was proposed based on sequence analysis (Presti et al., 2009). The orthomyxoviruses can infect a wide range of hosts. The Influenza A virus infects many mammalian (including human) and avian species, and it is capable of causing flu epidemics and pandemics that can cause a large number of deaths (Wang and Palese, 2009). The Influenza B virus infects humans as well as seals (Osterhaus et al., 2000), whereas the influenza C virus infects humans and pigs (Guo et al., 1983). The thogotovirus can also infect mammals, and it is usually transmitted by ticks (Jones and Nuttall, 1989). The Isavirus infects Atlantic salmon and other salmonid species (Raynard et al., 2001). In this thesis report, I will primarily focus on the influenza A virus and the Isavirus, although the other orthomyxoviruses may be very similar in terms of morphology, genome composition, virion structure, and replication cycle.
1.2 Influenza A virus

1.2.1 Overview

Influenza infections are a major public health concern each year, and the influenza A virus can cause widespread pandemics with a high mortality rate (Clark and Lynch, 2011; Lagace-Wiens et al., 2010). In order to better prevent and treat human influenza virus infections, it is critical that we obtain a comprehensive understanding of the basic molecular biology of influenza viruses.

Influenza A viruses, the causative agents of both epidemic and pandemic flu, are enveloped, single-stranded, negative-sense RNA viruses (Lamb and Krug, 2001). The influenza A virus has 8 linear single-stranded RNAs, encoding up to 12 viral proteins, including PB1, PB2, PB2-F1, N40, PA, NP, HA, NA, NS1, NEP, M1, and M2 (Medina...
and Garcia-Sastre, 2011; Palese and Shaw, 2007). Segment 2 encodes 3 proteins, the polymerase subunit polymerase basic protein 1 (PB1), a small protein PB1-F2, and an N-terminally truncated version of PB1, N40. Segment 7 encodes matrix protein 1 (M1) and matrix protein 2 (M2), and segment 8 encodes non-structure protein 1 (NS1) and nuclear export protein (NEP), both by the alternative splicing of the mRNAs. The structure of the influenza virus and a list of the important structural proteins are shown in Figure 1.1. The influenza A virus has 3 envelope proteins, the hemagglutinin (HA), the neuraminidase (NA) and the matrix protein 2 (M2, an ion channel protein), that are embedded in the lipid bi-layer of the viral envelope. This lipid bi-layer is derived from the host’s plasma membrane and it includes both cholesterol-enriched lipid rafts and non-raft lipids (Nayak et al., 2009; Scheiffele et al., 1999; Zhang et al., 2000). The HA and NA proteins preferably associate with the lipid rafts, whereas the M2 protein shows no preference (Nayak et al., 2009; Schroeder et al., 2005). The surface glycoproteins HA and NA are responsible for the regulation of cell entry and exit (Lamb and Krug, 2001; Palese and Shaw, 2007): HA helps in host recognition and facilitate the membrane fusion of the viral envelope with the endosomal membrane, by binding to the sialic acid-containing receptors of the host cell. The efficiency of HA binding is dependent on the type of sialic acid and linkage that connects the sialic acid residue to the oligosaccharide of the receptor. Human-infecting influenza viruses prefer to bind to oligosaccharides containing α2,6-linked N-acetylneuraminic acid, whereas avian-infecting influenza viruses prefer to bind to those containing N-acetylneuraminic acid with α2,3-galactose linkages, which explains why avian flu cannot cause human pandemics
based on human-to-human infection (Lamb and Krug, 2001; Palese and Shaw, 2007). The NA glycoprotein cleaves sialic-acid receptors from the host cell membrane to release the progeny virions after virus replication is complete. The M2 protein is a pH-controlled proton channel, which is activated by endosomal acidification to inject protons into the virus interior that leads to the dissociation of M1 from the ribonucleoprotein (Pinto and Lamb, 2006). M1 forms a layer underneath the viral envelope and connects with the viral genomes (Lamb and Krug, 2001). Inside the M1 inner layer are the eight single-stranded, negative-sense, genomic RNA molecules, encapsidated by many nucleoproteins (NP) to form double-helical, rod-shaped structures, the ribonucleoprotein (RNP) complexes (Portela and Digard, 2002), as shown in Figure 1.2. Each RNP is associated with a heterotrimeric RNA polymerase, consisting of the polymerase basic protein 1 (PB1), the polymerase basic protein 2 (PB2) and the polymerase acidic protein (PA). The three polymerase proteins are responsible for the transcription and replication of viral RNAs (Resa-Infante et al., 2011; Ruigrok et al., 2010). Some influenza A viruses express a pro-apoptotic protein PB1-F2 that is encoded by an alternative open-reading frame (ORF) on the PB1 gene segment (Palese and Shaw, 2007). The virus gene segment NS can encode two

**Figure 1.2** Cartoon model of RNP organization (Portela and Digard, 2002). The RNP forms a double helical structure with the 5’- and 3’- ends bound by a heterotrimeric polymerase. NP molecules are shown in purple circles and RNA is shown as the black curved line.
non-structural proteins (NS1 and NS2) by alternative splicing (Lamb and Krug, 2001). NS1 is able to inhibit the translation of cellular mRNA and stimulate the translation of viral mRNA whereas NS2 (also name NEP, nuclear export protein) possesses a nuclear export signal that is important for the export of the newly-synthesized RNPs from the nucleus to the cytoplasm.

1.2.2 Viral Life cycle

The life cycle of the influenza A virus is shown in Figure 1.3. During infection, the influenza A virus enters the host cell by clathrin-mediated endocytosis. Upon receptor recognition mediated by the HA protein, the virus enters the host cell in an endosome.

**Figure 1.3** Life cycle of the influenza A virus. The RNPs are represented by helical hairpins, with the polymerase subunits (red, brown, and green) and NP (cyan) shown in different colors. In the nucleus, the viral transcription and replication processes are depicted according to the model proposed by Jorba et al. (Jorba et al., 2009). The Figure is modified from Das et al. (Das et al., 2010).
The endosome has an acidic environment, which triggers a conformational change of the HA proteins, resulting in the fusion of the viral and endosomal membranes (Skehel and Wiley, 2000). The low-pH environment also opens the M2 ion channels, causing the influx of protons to the viral particle’s interior. This causes the RNPs to dissociate from the M1 proteins (Pinto and Lamb, 2006), and become available for nuclear import.

Influenza viruses are among the few RNA viruses that replicate in the nucleus. Viral RNPs enter the host’s nucleus by active transport (Boulo et al., 2007). In the nucleus, the RNPs from the infecting virus serve as active templates for the synthesis of viral mRNA as well as anti-genomic, complementary RNAs (cRNA). The cRNAs are replication intermediates that direct the synthesis of nascent virion RNAs (vRNAs) (Lamb and Krug, 2001). The transcription products, the viral mRNAs, are translated in the cytoplasm. Newly synthesized NP, PB1, PB2 and PA are imported into the nucleus for the assembly of new RNPs (Boulo et al., 2007). The M1 and NEP also diffuse into the nucleus to facilitate the export of the RNPs (Cros and Palese, 2003). For nuclear export of the RNPs, M1 connects the RNPs to the NEP, which has a nuclear export signal (NES) that facilitates the transportation of the RNPs to cytoplasm (Cros and Palese, 2003; Lamb and Krug, 2001). In the cytosol, influenza virus RNPs are transported to the cytoplasmic membrane where they are selectively packaged into budding virions (Hutchinson et al., 2010; Palese and Shaw, 2007).
1.3 The ribonucleoprotein (RNP) complex

As a complex genomic entity with a unique structure and function, the influenza A virus RNP has been a subject of extensive study since the 1960's. RNPs purified from virions have been studied by electron microscopy (EM) in great detail (Compans et al., 1972; Pons et al., 1969), and have been shown to possess a rod-shaped structure ~10nm in diameter. Compans et al. also found that the purified RNPs could be categorized into three length groups, possibly due to the different lengths of their associated genes (Compans et al., 1972). It was also determined via negative-staining EM that the rod-shaped RNPs are structurally flexible and appeared to adopt a right-hand, double-helical structure (Compans et al., 1972). The viral polymerase complex was found to bind to one end of the helical rod by immuno-EM (Murti et al., 1988). It is also worthwhile to note that the RNPs were able to maintain an intact structure even in the absence of vRNA, indicating that the NP plays the major role in maintaining the overall structural organization and stabilization of the RNP (Ruigrok and Baudin, 1995).

1.3.1 The components of the RNP

a. The genomic RNAs

The genome of the influenza A virus consists of eight segments of single-stranded, negative-sense RNAs, coding for up to 12 viral proteins as described earlier. Each segment of RNA serves as the template for the transcription of two types of positive-sense RNAs, the mRNA which is used for protein translation, and the cRNA which is an intermediate in the replication of the viral genome (Lamb and Krug, 2001;
Palese and Shaw, 2007). These two types of (+) RNAs have distinct sequences. The mRNAs are capped at their 5’-termini by a 10-13nt sequence snatched from host pre-mRNAs, and poly-adenylated at the 3’-end by repetitive copying of the oligo-U sequences near the 5’-termini of the vRNA templates. Different from the mRNAs, the cRNAs are neither capped nor poly-adenylated. The cRNAs are encapsidated by NPs like the vRNAs.

Each of the genomic RNA in the influenza A virus has a conserved, 12-13nt long sequence at both the 5’ and 3’ ends. These conserved 5’ and 3’-terminal sequences are partially complementary to each other, so that the viral RNA can “circularize” to form a “panhandle” structure through base-pairing (Hsu et al., 1987; Klumpp et al., 1997). The “panhandle” structure is important for RNP assembly and genomic packaging because it provides a binding site for the polymerase, which functions in both transcription and replication (Desselberger et al., 1980; Hagen et al., 1994; Luytjes et al., 1989; Neumann et al., 2004; Neumann and Hobom, 1995; Robertson, 1979; Skehel and Hay, 1978). The RNP will not assemble without an intact panhandle and the RNA will not be packaged into progeny virions even at the presence of a packaging signal (Luytjes et al., 1989; Neumann and Hobom, 1995). The complementarity of the 5’- and 3’-termini is partial, which may help to distinguish the cRNAs from the vRNAs to prevent the packaging of the former into the progeny virions (Neumann et al., 2004). Specifically, the central unpaired A10 at the 5’ end of each vRNA is critical for this purpose (Tchatalbachev et al., 2001).
The RNPs of the influenza A virus adopt a unique structure compared to the nucleocapsids from other non-segmented, negative-sense RNA viruses. The influenza A virus RNPs unwind under high/low salt conditions, giving rise to closed circular structures (Klumpp et al., 1997). Chemical probing experiments indicate that vRNA binds to the NPs with its phosphate backbone and that the nucleotide bases are exposed to solvent (Elton et al., 1999). The vRNAs associated with RNPs can be replaced by polyvinylsulfate (PVS), a negatively charged polymer (Goldstein and Pons, 1970), which also suggests that the influenza NP-RNA interaction is via the negatively-charged backbone instead of the bases. Furthermore, the vRNAs in the influenza virus RNPs were readily digested by RNase treatment (Baudin et al., 1994), suggesting that very little protection was provided by the NP encapsidation.

Besides the primary sequences of the genomic RNAs, RNA structure, at both the secondary and tertiary levels, are also important for virus replication. Studies on the RNA secondary structures of the influenza A virus have been primarily focused on the NS segment (Gultyaev et al., 2007; Ilyinskii et al., 2009). It was proposed that both secondary structures and tertiary structures are important for alternative mRNA splicing of the NS segment (Nemeroff et al., 1992; Plotch and Krug, 1986). The RNA structures have also been proposed to be important in the selective packaging of the eight RNPs. For instance, RNA packaging signals may form short secondary structures that help to link eight different RNPs together (possible by non-canonical base-pairing) (Fujii et al., 2003; Hutchinson et al., 2008; Liang et al., 2008; Marsh et al., 2008; Muramoto et al.,
At present more evidence is needed to support this packaging model, especially the identifications of specific recognition sites. Recently, a bioinformatics study on all influenza A coding regions has identified 20 potential conserved RNA secondary structures, which may cover those signal sites important for RNP packaging (Moss et al., 2011).

b. The nucleoprotein (NP)

**Figure 1.4** The NP structures of the influenza viruses. **A.** The crystal structure of the influenza A virus NP. The influenza A virus NP forms trimer in the crystal. Each NP molecule has a head domain and body domain, with the potential RNA-binding groove located in between (Ye et al., 2006). **B.** The crystal structure of the influenza B virus NP. The influenza B virus NP also has a head domain and body domain, whose structure resembles those of the influenza A virus NP (Ng et al., 2012b).

The influenza A virus NP is a multifunctional protein that has been shown to interact with a number of viral (e.g., PB1, PB2, M1, etc) and host proteins (e.g., RAF-2p48/UAP56 and Tat-SF1, etc) (Portela and Digard, 2002; Prokudina et al., 2001). One of the NP's primary functions is to coat viral RNA to facilitate its folding into a double-helical RNP structure. To date, there are three published structures of influenza
NPs at the atomic level: two of the influenza A viruses and one of the influenza B viruses (Ng et al., 2012b; Ng et al., 2008; Ye et al., 2006). All three of the NPs with known structures formed ring structures in the absence of RNA. Although their oligomers vary in size, all three NPs assume the overall shape of a crescent with a head and a body domain. In between the two domains is a deep groove. This groove is postulated to function as the RNA-binding groove, as its surface is rich in basic amino acid residues that are likely to interact with the RNA phospho-diester backbone (Elton et al., 1999; Ye et al., 2006). For the influenza viruses, 11 out of the 16 basic amino acid residues from this groove are conserved between the influenza A, B, and C viruses, which supports the idea that this grove is involved in RNA binding (Ye et al., 2006). It has been shown that the mutation of four arginine residues from the two flexible loops within the groove resulted in a dramatic reduction in the RNA binding affinity of influenza A virus NP (Ng et al., 2008). The aromatic residue Y148, located at one end of the groove, is also conserved and believed to stabilize RNA by base stacking (Ye et al., 2006).

Oligomerization of the NP is mediated by an extended tail-loop structure (i.e. aa402 – 428 in the influenza A virus NP) located at the back of each NP molecule (Ng et al., 2008; Ye et al., 2006). The tail-loop maintains the NP-NP interaction by inserting itself into the binding pocket of the neighboring NP molecule. The interaction involves various types of intermolecular forces: intermolecular β-sheets, hydrophobic interactions, and salt bridges (Ye et al., 2006). The interaction mediated by the tail loop has been
shown to be required for NP oligomerization, because a deletion mutant without the tail loop forms only monomers in solution (Ye et al., 2006). Furthermore, gel filtration analysis and EM studies showed that mutating either of the two residues making an inter-molecular salt bridge (i.e. R416 and E339) also eliminated NP oligomerization (Ye et al., 2006). The loop-binding pocket could be a good target for small-molecule drugs mimicking the tail loop structure, such that competitive binding of the drug could disrupt NP oligomerization and thus interfere with viral replication.

c. The polymerase complex (PA, PB1, PB2)

The influenza virus polymerase is a heterotrimeric complex consisting of PA, PB1 and PB2, with multiple enzymatic and ligand binding activities necessary for the synthesis of capped, poly-adenylated mRNAs during transcription as well as full-length genomic/anti-genomic RNAs during replication (Lamb and Krug, 2001). The 3D structure of the heterotrimeric polymerase complex has been studied by cryo-electron-microscopy. The polymerase samples were obtained from different sources, including recombinant RNP, polymerase-vRNA complexes produced by in vivo replication, as well as the free soluble polymerase devoid of template (Area et al., 2004; Coloma et al., 2009; Resa-Infante et al., 2010; Torreira et al., 2007). As the core RNA-dependent RNA-polymerase (RdRp) subunit, PB1 interacts with the C-terminal domain of PA (PAc) by its PB1N domain, and binds to the N-terminal domain of PB2 (PB2N) with its PB1C domain. Despite the relatively-low resolution of the model, it was found that the morphology of the free soluble polymerase complex is different from that
when it is within an RNP or with a vRNA bound, which suggests that the polymerase complex undergoes major conformational changes when interacting with other viral or cellular components (Area et al., 2004; Coloma et al., 2009; Resa-Infante et al., 2010; Torreira et al., 2007).

The influenza polymerase complex has also been studied by X-ray crystallography (Figure 1.5). So far, the known crystal structures include: the 25 kD N-terminal PA domain which displays the endonuclease activity needed for cap-snatching (Dias et al., 2009; Yuan et al., 2009), the 55 kD C-terminal PA domain that mediates the PA-PB1 interaction (Guu et al., 2008; Hara et al., 2006), the PB2 aa318-483 domain that binds to the 5’ pre-mRNA cap (Guilligay et al., 2008), the PB2 aa538-676 domain (the 627-domain) involved in host adaptation (Tarendeau et al., 2008), and the PB2 C-terminal NLS-domain (aa686-757) that binds cellular importins (Tarendeau et al., 2007). PB1, the largest subunit of the polymerase, hosts the polymerase catalytic active site (Biswa and Nayak, 1994; Kobayashi et al., 1996; Poch et al., 1989) as well as specific binding sites for the conserved 5’ and 3’-vRNA termini (Fodor et al., 1993). To date, the only known structures of PB1 are the N-terminal fragment (aa1-25) (He et al., 2008; Obayashi et al., 2008) which interacts with PA, and the C-terminal fragment which forms a three helical bundle (aa678-757) that interacts with the N-terminus of PB2 (Sugiyama et al., 2009).
Figure 1.5 The crystal structures of the influenza A virus polymerase complex. A. 2.2Å crystal structure of the N-terminal 197 residues of PA, in complex with a bound magnesium ion shown as a silver sphere (Yuan et al., 2009). B. The crystal structure of the PA\textsubscript{C} domain in complex with the short PB\textsubscript{N} \(\alpha\)-helix. The PA helices are colored in red, PA strands are colored in yellow, and PA coil is colored in green. The PB1 residues are in dark blue (Obayashi et al., 2008). C. The crystal structure of PB1\textsubscript{C}-PB2\textsubscript{N} complex. Helices from PB1 are colored in red, and helices from PB2 are colored in blue. Coil regions are colored green (Sugiyama et al., 2009). D. The crystal structure of the PB2 cap-binding domain (aa318-483) in complex with m\(^7\)GTP. Helices are in red and strands are in yellow. The m\(^7\)GTP is shown as a ball-and-stick model (Guilligay et al., 2008). E. The crystal structure of the PB2 627-domain (aa538-676). Helices are shown in red and strands are colored in yellow (Tarendeau et al., 2008). F. The crystal structure of the PB2 C-terminal domain (aa686-757, red), in complex with human importin 5 (blue), comprising ten armadillo repeats (Tarendeau et al., 2007).
The PA\(_N\) domain has an endonuclease activity that is dependent on divalent cations. In the crystal structure reported by Yuan et al, a magnesium ion was found to interact with conserved residues in the catalytic cavity (Yuan et al., 2009). Mutating these important residues eliminated the cap-snatching activity of the polymerase but did not affect genome replication. Instead of a magnesium ion, two manganese ions were found inside the PA acidic cavity by Dias et al. One of these two manganese ions occupies the same location as the Mg\(^{2+}\), and the other Mn\(^{2+}\) is 4Å away, suggesting a two-metal catalytic mechanism (Dias et al., 2009). The PA\(_C\) domain interacts with the N-terminal 14 residues of the PB1 protein according to the structure of the PA\(_C\):PB1\(_N\) complex (He et al., 2008; Obayashi et al., 2008). The N-terminal residues of PB1 forms a short \(\alpha\)-helix, which interacts with a hydrophobic pocket on PA\(_C\) that is surrounded by secondary structure elements \(\alpha8, \alpha10, \alpha11, \alpha13, \beta8,\) and \(\beta9\). Residues on the PA-PB1 interface are highly conserved among different isolates of influenza viruses.

There are three PB2 fragments that have had their structures determined. Residues aa318-483 fold independently into a domain that binds to the 5’ caps of host pre-mRNAs for cap-snatching mediated by PA. This domain was co-crystallized with m\(^7\)GTP (Guilligay et al., 2008), the guanine base of which was stacked with the aromatic ring of Phe404. The N1 and N2 atoms of the guanine form hydrogen bonds with the acidic residues Glu361, and Lys376 that are also involved in base recognition by interacting with the guanine atom O6. Although the PB2 cap-binding domain has a unique structural fold, the chemical coordination used by PB2 for the recognition of the
methylated base is similar to other cap-binding proteins previously characterized, including eIF4E (Marcotrigiano et al., 1997), the nuclear cap binding complex (CBC) (Mazza et al., 2002), and the vaccinia virus VP39 (Hodel et al., 1998). In these proteins, the m'G is usually sandwiched by two aromatic residues, one of which is often a tyrosine or tryptophan. Furthermore, there is usually an acidic residue nearby to neutralize the positive charge due to N7 methylation. The PB2 627-domain containing residues aa538-676 is enriched with amino acids involved in host adaptation, including the residue 627 that is mutated from Glu to Lys when avian viruses adapt to human hosts (Tarendeau et al., 2008). The crystal structure of this domain has a previously unobserved fold with the Lys627 exposed on the surface. The K627E mutation did not induce a structural change, but the resulting charge reversal could have disrupted a basic surface patch and thus possibly affected its RNA-binding, which could be involved in the adaption of the virus to different host (Tarendeau et al., 2008). This 627-domain was found to be involved in the PB2-NP interaction, and the strength of this interaction has been correlated with the polymerase activity (Ng et al., 2012a). The last PB2 domain with a determined structure is the C-terminal NLS-domain (aa686-757) which contains a classical bi-partite nuclear localization sequence K736RKR739X12K752RIR755 (Tarendeau et al., 2007). Tarendeau et al. determined the solution structure of the NLS domain by NMR and the complex structure of this domain with importin-α5 by X-ray crystallography (Tarendeau et al., 2007). After fusing the NLS domain with an eGFP reporter, it was observed to accumulate in the nucleus, whereas the nuclear localization was abolished by mutations to the NLS sequence. By comparing the holo- and apo- structures of this
NLS-domain, it was noted that the region 736–759 unfolds to allow importin to bind to the bipartite NLS (Tarendeau et al., 2007). In addition, Asp701 of the PB2 NLS domain was found to be important for the acquisition of pathogenicity as avian viruses adapted to mammalian species, possibly because it forms a salt bridge with Arg753, another important residue in the NLS (Tarendeau et al., 2007).

The PB1 subunit is the largest subunit in the polymerase, and it contains the RdRp catalytic active site (Biswa and Nayak, 1994; Kobayashi et al., 1996; Poch et al., 1989). To date, there is no available atomic-level structure of the influenza RdRp active site. Besides the N-terminal fragment which interacts with PA, the other determined crystal structure of PB1 is that of a C-terminal three helical bundle (aa678–757) in complex with the N-terminal 35 residues of the PB2 subunit (Sugiyama et al., 2009). A number of residues at the interface, especially Val715 in PB1, were shown to be important for viral RNA synthesis (Sugiyama et al., 2009). Mutating these residues inhibited RNA synthesis, potentially by disrupting the assembly of the polymerase complex (Sugiyama et al., 2009). To understand how the PB1 subunit functions as an RdRp, a high resolution structure and corresponding functional studies are needed.
1.3.2 The RNP Structure

The first three-dimensional structure of the influenza A virus RNP is that of an artificial mini-RNP (Area et al., 2004; Coloma et al., 2009; Martin-Benito et al., 2001; Ortega et al., 2000). To circumvent the structural flexibility problem, a mini-RNP was generated from *in vivo* amplification by expressing the three polymerase subunits, the NPs, and a 248nt model vRNA containing the highly conserved terminal sequences (Martin-Benito et al., 2001). The mini-RNP is more structurally rigid than the native RNPs, thus allowing cryo-EM reconstruction to ~12Å resolution (Figure 1.6 A). Cryo-EM reconstruction of the mini-RNP shows a closed ring structure consisting of nine NP molecules, with a copy of the viral polymerase attached to the outer edge of the otherwise symmetric ring (Coloma et al., 2009). The viral polymerase adopts a compact shape and simultaneously interacts with two adjacent NP molecules (Coloma et al., 2009). Each NP molecule shows two-domain morphology that agrees with previously determined NP crystal structures (Ng et al., 2008; Ye et al., 2006). vRNA cannot be readily discerned at this resolution, but it presumably constitutes some of the adjoining densities that connect neighboring NP molecules. Due to the limited resolution, the boundaries between the three polymerase subunits were not obvious. Using engineered fusion tags and monoclonal antibodies, Area et al., was able to map the rough location of PA, PB1, and PB2 in the polymerase complex (Area et al., 2004). It was found that the polymerase contacts with two NP monomers via the PB1 and PB2 subunits (Coloma et al., 2009; Martin-Benito et al., 2001), consistent with previous biochemical studies (Biswas et al., 1998; Mena et al., 1999). The RNP-associated polymerase shows similarities in overall
structure compared to the EM reconstruction of a free polymerase (Area et al., 2004; Torreira et al., 2007), but it is also clear that some conformational changes have taken place upon interacting with the NP and/or the vRNA template.

**Figure 1.6** The influenza A virus RNP structure. **A.** Cryo-EM reconstruction of the mini-RNP (Coloma et al., 2009; Martin-Benito et al., 2001). The NP crystal structure (Ye et al., 2006) is fitted into the electron density. The arrow points to adjoining densities that are presumably made of vRNA and the NP oligomerization tail loop. **B.** Cryo-EM reconstruction of a double-helical RNP by Arranz et al., (Arranz et al., 2012). The viral polymerase complex is located at the bottom end of the RNP and is shown in green and orange. The two opposite-running NP-RNA strands are colored differently in blue and pink. The NP-RNA turning loop on the top end of the RNP is highlighted in dark green. **C.** Cryo-EM reconstruction of a RNP by Moeller et al. (Moeller et al., 2012). On the right is a model showing the RNP organization. The viral polymerase is highlighted in red. **D.** The helical stem of the RNP from (B) is fitted with the NP crystal structure and modeled with RNA (in yellow) (Arranz et al., 2012). **E.** Central filament region from (C) fitted with the NP crystal structure protomers. Arrows indicate RNA polarity (Moeller et al., 2012).
In an exciting development, cryo-EM reconstructions of authentic RNPs were reported by two research groups recently (Arranz et al., 2012; Moeller et al., 2012) (Figure 1.6 B-E). The RNP reconstruction reported by Moeller et al., used RNPs generated by *in vitro* expression of the four RNP proteins (*i.e.* PA, PB1, PB2 and NP) via transient transfection of a human cell line in the presence of their respective vRNA segments (Moeller et al., 2012). At ~20 Å resolution, the final model confirms that the RNP adopts a double helical structure with two anti-parallel strands leading to and away from the polymerase that is located at one end of the RNP (Figure 1.6 B, D). The double-helical stem region shows a rise between two neighboring NP of 32.6 Å with 4.9 NP molecules per turn. The other cryo-reconstruction of the influenza A virus RNP was reported by Arranz et al., using native RNPs purified from virions (Arranz et al., 2012). The structure also shows a double-helical stem with major and minor grooves (Figure 1.6 C, E). The rise step between adjacent NPs is 28.4 Å with a rotational angle of 60° and 6 NPs per turn on each strand (Arranz et al., 2012). In both RNP structures, the putative RNA binding groove of the NP scaffold is exposed on the outer surface of the RNP. Assuming that the positively charged groove of the NP serves as the RNA binding site, a vRNA was built into the final model. By following a contoured path, ~120-150 nucleotides of RNA are placed in each helical turn.

It is also important to note that the two models by Moeller et al., and Arranz et al., exhibit significant variations in helical parameters and NP orientations (Arranz et al.,
Using crystal structure docking, Arranz et al., proposes that the NP molecules from opposing strands contact each other through their body domains at a region near the disordered N-terminus of the NP structure. Moeller et al., however, suggests that the RNP helix is stabilized by the NP-RNA strand interacting with the opposing strand near the NP head domains. The most likely cause of the model difference is the different handedness of the two reconstructions, with Arranz et al., showing a left-handed helix and Moeller et al., showing a right-handed helix (Figure 1.6 D, E). It is expected that the modest resolution, rotational freedom of the NP molecules, and the source of the RNP samples (viral particles vs. cells) may also contribute to some inter-model variations as well. It is worth mentioning that Ye et al., recently reported a NP dimer crystal structure with a dimer interface that does not involve the tail loop (Ye et al., 2012). Mutational analysis indicated that the dimer interface is biologically relevant, suggesting a possible role in RNP assembly. Comparing the NP dimer structure with the RNP reconstructions may help to interpret interactions made between the two opposing NP-RNA strands.

The two cryo EM reconstructions of the RNP also offer a new look at the viral polymerase. Both Arranz et al., and Moeller et al., located the viral polymerase at the open end of the RNP hairpin, simultaneously interacting with both the 5' and 3'-ends of the vRNA (Arranz et al., 2012; Moeller et al., 2012). The close end of the RNP hairpin contains a small loop formed by a curved array of three to eight NP molecules. It was proposed by Moeller et al. that the PA C-terminal domain is structurally flexible and may
help to feed the vRNA template into the polymerase active site, based on structural homology between the PA C-terminal domain and the N-terminal domain of the Reovirus RNA polymerase (He et al., 2008; Moeller et al., 2012; Tao et al., 2002). Arranz et al. observed that the RNP-associated polymerase samples adopt two alternative conformations, but higher resolution structural information is needed to address the biological relevance of this distribution and the possible implications (Arranz et al., 2012).

### 1.3.3 RNP in viral transcription and replication

The polymerase complex of the influenza A virus is the core machinery for viral RNA replication and transcription (Lamb and Krug, 2001; Palese and Shaw, 2007). Viral RNA replication is primer independent, but transcription initiation requires short, 10-13nts long, capped RNA fragments snatched from host pre-mRNAs. The termination process for viral RNA replication and transcription is also different. Viral RNA transcription prematurely terminate at a polyU tract ~25 nts away from the end of the viral template, at which point the polymerase engages in repetitive copying of the polyU sequence for the synthesis of a polyA tail. In contrast, during replication the viral polymerase is able to read through the polyU, resulting in a faithful copy of the entire template.

The new RNP structures provide insight into our understanding of influenza virus transcription and replication (Arranz et al., 2012; Moeller et al., 2012). As the RNP
structure is predominantly maintained by the NP and the bound RNA is fully solvent exposed, the viral polymerase moving along the RNA template during viral RNA synthesis should result in little or only local disruption of the double-helical hairpin structure. During infection, viral mRNAs were detected immediately, but cRNAs were detectable only after viral protein synthesis started (Hay et al., 1977). Viral RNA replication requires soluble NPs for the elongation of nascent RNA chains, as a polymerase-RNA complex can only synthesize small-sized RNAs if the NP is not present (Honda et al., 1987; Honda et al., 1988; Resa-Infante et al., 2010). Several models, including the template modification model, the polymerase modification model, and the stabilization model, have been proposed to explain the NP’s role in replication and how the transcription and replication activities of the viral polymerase appear to occur during different phases of virus infection (Resa-Infante et al., 2011). These models, which are not necessarily mutually exclusive, entail the NP binding to the viral polymerase, as well as the viral template RNA and product RNA, respectively.

Recently, it has been demonstrated that the polymerase molecules that catalyze transcription and replication are of different origins (Jorba et al., 2009). By performing an in vivo complementation assay using mutated polymerase defective in either replication or transcription, Jorba et al. proposed a model in which the transcription takes place in cis-via the same polymerase that is part of the RNP; whereas the replication occurs in trans-with exogenous polymerases synthesizing the RNA and mediating the assembly of
progeny RNP, as shown in Figure 1.7 (Jorba et al., 2009). According to the model, viral replication starts with the binding of a second polymerase to the polymerase on the template, with this binding, allowing the second polymerase to access the 3’- end of the template RNA to initiate RNA synthesis. As the new RNA is synthesized, a third polymerase, or a second exogenous polymerase, binds to and protects the 5’- end of the new RNA, and possibly recruits NPs for the assembly of the new RNP. However, contradictory evidence by Vreede et al. showed that cRNA can accumulate in the beginning of infection if the NP and a catalytically-inactive polymerase are provided to protect the RNA from degradation, suggesting that replication can also occur in cis (Vreede et al., 2004). Therefore, more evidence is needed to elucidate the mechanism that influenza viruses use to replicate their genomes.

**Figure 1.7** Model of the RNP replication and transcription for the influenza A virus (Jorba et al., 2009). **A.** The replication steps. The replication occurs *in trans* with a second, exogenous polymerase synthesizing the RNA and a third, exogenous polymerase stabilizing the free 5’ ends. Multiple replications can take place simultaneously on the same RNA template. **B.** The transcription steps. The transcription takes place *in cis* where the same polymerase on the RNP transcribes the vRNA to mRNA. In the end, polymerase stutters around the oligo-U sequence to generate the polyA tail due to steric hindrance.
1.3.4 The RNP assembly

Newly synthesized NPs are imported back into the host cell nucleus to promote viral RNA replication and RNP assembly (Davey et al., 1985; Honda et al., 1988). As the NP binds to RNA non-specifically and has a strong tendency to self-polymerize, it is important to keep the NP in a soluble, encapsidation-competent state prior to RNP assembly. Unlike many other non-segmented, negative-sense RNA viruses, the influenza A virus does not encode viral proteins that are known to interact with the NP and prevent its self-oligomerization. It has been proposed that phosphorylation may play an important role in regulating the self-polymerization and RNA binding activities of the influenza A virus NP (Arrese and Portela, 1996; Tao and Ye, 2010; Yamanaka et al., 1990). Additionally, a number of host factors have been found to be important for influenza virus replication and RNP assembly (Kawaguchi et al., 2011; Momose et al., 2001; Momose et al., 1996; Momose et al., 2002; Naito et al., 2007). For example, RAF-2p48/UAP56 and Tat-SF1 assist in the formation of the vRNA-NP complex, possibly by functioning as chaperones to suppress the nonspecific aggregation of NPs (Momose et al., 2001; Naito et al., 2007). Interestingly, Ye et al., recently reported that the self-oligomerization activity of the NP is weak in the absence of RNA, but the interaction is kinetically stable once the NP oligomerizes (Ye et al., 2012). Therefore, it is possible that NPs remain monomeric until they encounter v/c RNAs in the cell nucleus.

RNP assembly in host cells requires only four viral proteins: NP, PA, PB1 and PB2. As shown by Moeller et al., in vivo amplification of vRNAs with only these four proteins
co-expressed in transfected cells produced rod-shaped RNPs with regular helical symmetry (Moeller et al., 2012). Therefore, although M1 and NEP are needed for RNP export from the nucleus (see below), they do not play major role in either the assembly or structural maintenance of the RNP. This feature is again different from other non-segmented, negative-strand RNA viruses (i.e. rhabdoviruses), in which the matrix protein plays a major role in organizing the helical nucleocapsids by simultaneously interacting with NPs from two adjacent helical turns (Ge et al., 2010).

It is likely that the NP initiates viral RNA replication via interactions with the polymerase (possibly PB1 or PB2), during which process the RNA-binding ability of NP is not required (Biswas et al., 1998; Mena et al., 1999; Newcomb et al., 2009). Therefore, the NP-polymerase interaction should facilitate the initial recruitment of NPs to newly synthesized vRNAs. Further vRNA encapsidation is likely stimulated by cooperative NP-RNA interactions (Yamanaka et al., 1990). Tarus et al., reported that in vitro NP oligomerization is a slow process that depends on the RNA length, with the oligomerization rate increasing drastically as the RNA length increases (Tarus et al., 2012). On average, each NP is associated with ~22-28 nts of RNA in the RNPs (Martin-Benito et al., 2001; Ortega et al., 2000). The viral polymerase is essential for maintaining the super-coiled RNP structure, since single-stranded RNA was observed when the polymerase was removed from the RNP (Klumpp et al., 1997).
It is unclear when the polymerase-NP-RNA complex collapses into the super-coiled, double helical RNP structure during replication. The fact that mini-RNPs consisting of nine NP molecules only form circularly shaped rings suggests that the condensation of NP-RNA polymers into double-helical structures does not occur until nascent RNPs reach certain sizes. Moeller et al. reported the observation of “branched” RNPs and suggested that the branches are made of partially replicated, nascent RNPs budding from the full-length RNP templates (Moeller et al., 2012). Using immuno-labeling, they showed that a second copy of the polymerase is located at the branching site in some RNPs. The observation of a second polymerase molecule is consistent with the notion that template RNPs are replicated in trans by free polymerase complexes and not by the polymerase molecule bound to the parental RNP (Jorba et al., 2009). Although “budding” RNPs sounds like an attractive interpretation of these branched RNP structures, whether they are truly replication intermediates or perhaps mis-folded RNPs still awaits further verification. One potential concern is that the length of the budding RNPs does not seem to correlate with the location of the RNP branches (Moeller et al., 2012).
1.3.5 The RNP trafficking

After the onset of infection, RNPs released from the infecting influenza A viruses are actively transported from the cytosol into the nucleus. It is not clear whether all eight RNPs are imported to the nucleus as a large bundle, or if they are separately imported as individual RNPs. All RNP component proteins contain at least one nuclear localization signal (NLS) that is necessary for nuclear import. Two regions of the PA protein, residues 124-139, and 186-247, were found to contain NLSs (Nieto et al., 1994). For PB1, a NLS was first found between residues 187-211 (Nath and Nayak, 1990), but later findings showed that the co-expression of PA was important for the efficient nuclear import of PB1 (Fodor and Smith, 2004). PB2 has a linear NLS with a sequence of K\textsubscript{736}RKR\textsubscript{739} that was shown to interact with importin-α in a co-crystal structure (Mukaigawa and Nayak, 1991; Tarendeau et al., 2007). Two NLSs have been identified in the NP sequence. One of these was a classical bipartite NLS that was found between residue 198 and 216 with a sequence of K\textsubscript{198}RX\textsubscript{199}RKTR\textsubscript{216} (Weber et al., 1998). A non-conventional NLS (nNLS) with a consensus sequence of S\textsubscript{3}QGTKRSYXXM\textsubscript{13} was also identified at the N-terminus of the NP (Neumann et al., 1997; O'Neill et al., 1995; Wang et al., 1997). Although all component proteins of the RNP carry their own NLSs, the NP is the major contributor for RNP import (O'Neill et al., 1995; Wang et al., 1997; Wu et al., 2007a). By dot blotting and immuno-gold labeling vRNPs, Wu et al. showed that the nNLS of the NP was much more accessible than the classical bipartite NLS, and that the labeled gold particles showed a regular periodicity which suggested a regular helical conformation of the RNP (Wu et al., 2007b). In addition, Cros et al. reported that
mutations in the nNLS completely abolished NP import, and that short peptides mimicking the nNLS competitively inhibited the nuclear import of the RNP (Cros et al., 2005). These findings are consistent with the crystal structures of the influenza NPs (Ng et al., 2008; Ye et al., 2006), as the nNLS is solvent exposed, structurally disordered, and can be easily fitted into the substrate binding pocket of importin-α.

Because influenza A virus assembly occurs at the host cell membrane, newly synthesized RNPs need to be exported out of the nucleus, thus travel in the opposite direction compared to their parental RNPs. No nuclear export signal (NES) has been found in the component proteins of the RNP. For RNP export, two other influenza proteins, M1 and NEP, are required as shown in Figure 1.3 (Bui et al., 2000; Neumann et al., 2000; O’Neill et al., 1998). The C-terminal domain of M1 is responsible for interacting with the RNP (Baudin et al., 2001), and M1-binding to the RNP likely helps to mask the NLSs on the RNP. Meanwhile, M1 can directly interact with the viral protein NEP which possesses a NES signal (O’Neill et al., 1998). The RNP-M1-NEP complex is recognized by the chromosome region maintenance 1 (CRM1) protein, which mediates the nuclear export of NES-containing proteins/complexes from the nucleus (Fukuda et al., 1997). The nuclear export of RNP has also been shown to require the viral activation of the cellular Raf/MEK/ERK (mitogen-activated protein kinase (MAPK)) signaling cascade that is activated late in the infection cycle, as blocking the cascade resulted in the retardation of RNP export and reduced titers of progeny virus (Pleschka et al., 2001). It has been shown that membrane accumulation of the influenza A virus hemagglutinin triggers the
activation of the MAPK cascade and induces RNP export. This may represent an auto-regulative mechanism that coordinates the timing of RNP export with virus budding (Marjuki et al., 2006).

1.3.6 Specific RNP packaging

Early evidence for selective packaging came from the use of defective-interfering influenza RNAs (DI RNAs) (Duhaut and McCauley, 1996; Odagiri and Tashiro, 1997). DI RNAs were shown to interfere with the incorporation of some specific gene segments while sparing others, suggesting that each segment contains a unique packaging signal. Later, further evidence was provided by reverse genetics, which revealed that all eight gene segments possess packaging signals that are required for efficient virion incorporation (Fujii et al., 2005; Fujii et al., 2003; Hutchinson et al., 2008; Hutchinson et al., 2010; Liang et al., 2005; Liang et al., 2008; Muramoto et al., 2006; Ozawa et al., 2007; Watanabe et al., 2003). EM studies have also provided strong evidence for selective packaging (Figure 1.8). The observation of the distinctive “7+1” pattern of the eight RNPs suggests that specific inter-RNP interactions maintain such a conformation (Noda et al., 2006) (Figure 1.8 A). Recent evidence by electron tomography showed that the RNPs of the “7+1” bundle are actually different, with four longer RNPs and four shorter RNPs of significantly different lengths, consistent with the length distribution of the eight influenza RNA segments (Fournier et al., 2012; Noda et al., 2012) (Figure 1.8 B). Electron tomography studies also revealed that the eight RNPs are aligned at the budding tip and interconnect with each other to form a supra-molecular assembly.
Figure 1.8 Specific packaging of the influenza A virus RNP. A. Linkages among the eight RNPs in a virion (Noda et al., 2012). A 0.5 nm-thick tomogram is shown with the eight RNPs highlighted in different colors. Short string-like structures can be seen between the RNPs (arrowheads). B. 3-D model of a multi-segmented RNP complex (Noda et al., 2012). The four long RNPs are shown in red, while the four shorter ones are shown in grey. C. Budding (left, middle) and mature (right) virions (Noda et al., 2012). Red curves in the left and middle columns indicate the membrane region where viral spike proteins are present. Shown below are schematic diagrams of the RNP packaging process Scale bar, 100 nm. A-C are taken from Noda et al. (Noda et al., 2012).

(Figure 1.8 C). Using fluorescence in situ hybridization (FISH) analysis at the single-virus particle level, Chou et al., confirmed that the eight unique RNPs are incorporated into
progeny virions by a selective packaging mechanism (Chou et al., 2012). Co-localization tests demonstrated that most of the virus particles have incorporated at least one copy of each of the eight RNPs. The exact copy number of each RNP was determined by comparing the photo-bleaching profiles of probes against the HA RNA segment (i.e. the RNA segment encoding the hemagglutinin protein) for a wild-type and a recombinant virus carrying two copies of the HA segments. Their results demonstrated that most virus particles contain only one copy of each of the eight RNP complexes.

The sequence-specific signals for influenza genomic packaging have been discovered for each of the eight genomic RNAs (Hutchinson et al., 2010). The “signal regions” cover the untranslated regions (UTRs) of both termini as well as the adjacent coding sequences of the open reading frame (ORF). Many approaches have been employed to map the regions containing the packaging signals. The earliest results were obtained from experiments on DI RNAs (Duhaut and Dimmock, 2000; Duhaut and Dimmock, 1998; Duhaut and McCauley, 1996; Hughes et al., 2000; Jennings et al., 1983; Nayak et al., 1982; Noble and Dimmock, 1995), which are shorter RNAs derived from the wild-type RNAs with certain region(s) deleted, while still maintaining the ability to be replicated and packaged. Examining these DI RNAs showed that the UTRs as well as the terminal coding sequences are well preserved in the DI RNAs, suggesting that they play a significant role in genomic packaging. Later, reverse genetics experiments confirmed that all eight influenza vRNAs have bipartite packaging signals located at the 5’ and 3’ termini (Fujii et al., 2005; Fujii et al., 2003; Liang et al., 2005; Marsh et al., 2007; Muramoto et al.,
It was demonstrated that the UTRs together with the terminal coding regions can significantly enhance the packaging efficiency, suggesting that the coding sequences also contribute to genomic packaging (Fujii et al., 2003). In addition, codons at the terminal coding regions were found to have synonymous variation rates significantly lower than expected, indicating that the RNA primary sequence is important and thus selectively preserved (Gog et al., 2007). Indeed, synonymous nucleotide mutations within the packaging signal regions produce recombinant viruses with reduced replication efficiencies (Fujii et al., 2005; Gog et al., 2007; Hutchinson et al., 2008; Hutchinson et al., 2009; Liang et al., 2008; Marsh et al., 2007; Marsh et al., 2008).

The localization of the packaging signals near the 5’ and 3’ termini of the vRNAs suggest that these RNA sequences should be mapped to the double-helical RNP s near the end where the viral polymerase is located. Using electron tomography 3D-reconstructions, Fournier et al. showed that the eight vRNPs contact each other at the budding tip of the influenza A virus particles (Fournier et al., 2012). This contact region is thick enough to accommodate all of the described packaging signals. They also demonstrated that in vitro all vRNAs are involved in a single interaction network, with each vRNA segment interacting with at least one other vRNA partner. Fournier et al. thus suggests that the RNP s are likely held together by direct base-pairings between packaging signals. This model raises an intriguing question as to how the packaging signals are presented on the surface of the RNP s. Is vRNA completely denatured by tightly wrapping around the NP
scaffold, or perhaps some NP-free RNAs exist in the RNP to allow vRNA-vRNA interactions? Although packaging signals are identified in all eight vRNAs, much still needs to be learned about the molecular details of their interaction.

1.4 Infectious Salmon Anemia Virus

1.4.1 Overview

The infectious salmon anemia virus (ISAV) is the first Orthomyxovirus that was detected in fish, and it is classified in the new genus *Isavirus* (Lamb and Krug, 2001; Palese and Shaw, 2007). It has been primarily observed in Atlantic salmon (*Salmo salar*), but occasionally in other salmonid species as well (Cottet et al., 2011). ISAV causes severe anemia in infected fish, often leading to wide-spread epidemics with high mortality rates (Cottet et al., 2011).

![Figure 1.9](image)
ISAV shares many common properties with the influenza virus. The morphology of the ISAV viral particle is similar to that of the influenza’s as shown in Figure 1.9 (Bouchard et al., 1999; Cottet et al., 2011; Kibenge et al., 2001). The ISAV also has eight segments of single-stranded, negative-sense RNAs in its genome, and each of the RNAs is coated with NPs, forming double-helical RNPs as shown in Figure 1.10 (Falk et al., 1997). EM images of the ISAV RNPs obtained from virions showed a double-helical structure with similar dimensions to those of influenza A virus. It was reported that the longest RNPs of ISAV were ~120 nm long, very close to the maximum length of ~110 nm measured for the influenza A virus RNPs (Compans et al., 1972; Falk et al., 1997). ISAV contains eight gene segments, with the longest one, which encodes PB2, containing ~2.4 kb of nucleotides. This is only slightly larger than the longest RNA segment of the influenza A virus.
virus, which is \(~2.3\) kb in length. Therefore, the ISAV and the influenza A virus have similarly sized RNPs as well as RNAs.

The eight ISAV genes can encode up to 11 proteins, most of which have counterparts in the influenza A virus with related functions, as discussed later in the following section. In addition, like influenza A virus, there is also a consensus sequence at each of the 5'-UTRs for all of the ISAV RNA segments that is complementary to the 3'-UTRs (Sandvik et al., 2000). This complementarity should allow the formation of a short dsRNA panhandle. In influenza viruses, such panhandles provide binding sites for the viral polymerase, which is important for the stability of viral RNPs as well as the viral transcription/replication process (Hsu et al., 1987; Klumpp et al., 1997; Sandvik et al., 2000).

Last but not least, the life cycle of ISAV (Figure 1.11) is very similar to the influenza life cycle as shown in Figure 1.3. The ISAV infection is initiated with sialic acid recognition by the ISAV HE, followed by endocytosis of the virus particles (Eliassen et al., 2000; Hellebo et al., 2004; Kristiansen et al., 2002; Workenhe et al., 2007). The acidic environment of the endosome triggers conformational changes to the F protein to induce the fusion of the viral and endosomal membranes. Meanwhile, protons are pumped through the M2 proton channel into the virus to release the RNPs (Aspehaug et al., 2005; Eliassen et al., 2000). The free RNPs are imported into the nucleus with the
help of the nuclear localization signals on the NPs and the polymerase proteins (Falk et al., 1997). The viral transcription and replication take place in the nucleus of the host cell. The mRNAs are exported and translated (Palese and Shaw, 2007; Sandvik et al., 2000). The surface glycoproteins F, HE, and M2 are translated by ribosomes associated with the ER, whereas the other viral proteins are translated in the cytoplasm (Falk et al., 2004; Palese and Shaw, 2007). The translated PB1, PB2, PA, and NP are actively transported into the nucleus for RNP assembly (Palese and Shaw, 2007). The M1 and NEP are imported into the nucleus by diffusion (Palese and Shaw, 2007). The vRNA replicates are assembled into RNPs and exported to the cytoplasm by M1 and NEP (Goic et al., 2008). Finally the RNPs are packed into progeny viruses that leave the host by budding (Palese and Shaw, 2007).

Figure 1.11 Schematic drawing of the ISAV life cycle. The proteins with question marks are those whose functions are assumed by comparison with the similar proteins of the influenza A virus. Image from Cottet et al. (Cottet et al., 2011).
1.4.2 ISAV proteins

The ISAV polymerase consists of the 79.5 kD PB2 (Snow et al., 2003), the 80.5kD PB1 (Krossoy et al., 1999), and the 65.3 kD PA (Ritchie et al., 2001), respectively. The ISAV PB2 is shown to be similar to the influenza PB2 in terms of the protein’s size, amino acid composition, and the presence of a bipartite nuclear localization signal (NLS) at the C-terminus, which is also found for all influenza PB2 proteins (Snow et al., 2003). PB2 has also been expressed in a fish cell line while fused with a GFP reporter protein to confirm its localization in the nucleus (Snow et al., 2003). The protein encoded by segment 2 was identified to be the ISAV PB1 based on the high sequence similarity between it and the influenza PB1 and the presence of the typical motifs A-E characteristic for RNA-dependent RNA-polymerases (Krossoy et al., 1999). The ISAV PB1 has a PI of 9.9, a molecular weight of 80.5kD, and a net charge of +22 at neutral pH, which are similar to other PB1 proteins in the family of the Orthomyxoviridae (Clouthier et al., 2002; Cottet et al., 2011). The ISAV gene segment 4 is found to encode a protein that is 579aa long with an acidic character. It is believed to be the ISAV PA protein based on its the size and pI in comparison to the influenza PA protein (Ritchie et al., 2001). Also, this ISAV PA was found to share 48% sequence similarity with part of the influenza C P2 polymerase protein (Ritchie et al., 2001).

Similar to influenza NPs, the ISAV-NP is a structural protein with RNA-binding activity (Aspehaug et al., 2004). Sequencing alignment analysis comparing the ISAV-NP’s amino acid sequence with existing databases did not find any close homolog, but the
PROPSEARCH program showed that its composition of the amino acids was similar to that of the influenza B NP (Cottet et al., 2011). It is known that the influenza NP binds to ssRNA non-specifically, but for the ISAV-NP, all that is known so far is that it could bind to ssRNA (Albo et al., 1995; Galarza et al., 1992). The putative NP was found to shuttle between the cytoplasm and the nuclei in ISAV-infected cells (Aspehaug et al., 2004). The NP is too big to be transported through the nuclear pore by diffusion, so active transport is needed. There are two mono-partite NLSs (Nuclear Localization Signals) first predicated by sequence analysis at $^{230}$RPKR$^{233}$ and $^{473}$KPKK$^{476}$ and later confirmed by site-directed mutagenesis (Aspehaug et al., 2004). Like the NPs of the influenza viruses, the ISAV-NP seems to be conserved among known ISAV isolates (Aspehaug et al., 2004; Clouthier et al., 2002; Ritchie et al., 2001; Snow and Cunningham, 2001). It was also reported that the C-terminal 30 amino acid residues of the ISAV-NP were extremely rich in acidic residues, similar to the NP of influenza viruses (Aspehaug et al., 2004; Portela and Digard, 2002).

Unlike other orthomyxoviruses, ISAV has two unique proteins, the F protein and the HE protein (Aspehaug et al., 2005; Krossoy et al., 2001a; Krossoy et al., 2001b). The fusion protein F is expressed by ISAV in the form of a ~50kD F0 precursor, which undergoes proteolytic cleavage to produce the 30kD F1 and 20kD F2 that are linked by disulfide bonds. The protease that performs the digestion is of extracellular origin, and the resulting F2 subunit exhibits fusion activity (Aspehaug et al., 2005). During viral infection, similar to the influenza HA protein, this fusion protein induces the fusion of the viral
membrane with the endosomal membrane, releasing the viral RNPs into the cytoplasm for nuclear import (Aspehaug et al., 2005; Eliassen et al., 2000). HE is a 42kD protein whose amino acid sequence does not show any similarity to other orthomyxovirus proteins. The 42kD protein was determined to be the ISAV hemagglutinin-esterase based on its ability to react with a monoclonal antibody (MAb) directed against the ISAV hemagglutinin (Krossoy et al., 2001a). As a surface glycoprotein, the influenza A virus HA is known to mediate host cell recognition and the subsequent membrane fusion (Eliassen et al., 2000). However, different from the influenza A virus HA, the ISAV HE has an additional acetylesterase activity to destroy host receptors. In contrast, influenza A viruses encode a separate protein NA, which destroys host receptors using neuraminidase activity (Palese and Shaw, 2007). Interestingly, influenza C viruses encode a HEF glycoprotein which is known to bind to cellular receptors containing 9-O-acetyl-5-N-acetylsialic acids and releases the acetyl residue from position C-9 of the substrate (Herrler & Klenk, 1987; Rogers et al., 1986).

Each of the two shortest RNA segments of ISAV can generate at least two mRNAs by alternative splicing (Palese and Shaw, 2007). Less research has been done on these two segments compared to the other RNA segments. The segment 7 ORF 1 may encode an ISAV non-structural protein or minor structural protein that is around 32kD (Bierin et al., 2002; Kibenge et al., 2007). A putative nuclear export protein (NEP) may be encoded by the 1/2 ORF of the segment 7, based on the observation that this protein has three leucine-rich motifs that are characteristic of the influenza NEP (Kibenge et al., 2007).
There may be a third protein encoded by this segment, but its function is not clear. Segment 8 can encode two proteins with estimated molecular masses of 22-24kD and 27.4kD, respectively (Bierin et al., 2002). The 22-24kD protein was assigned to be the Matrix protein (M1), and the 27.4kD protein may be a minor structural protein that is equivalent to M2 from the influenza A virus (Bierin et al., 2002; Falk et al., 2004; Kibenge et al., 2007).

1.4.3 ISAV can be a good model for studies of the influenza A virus

As described above, ISAV and influenza share many similarities in terms of virus structure, genomic composition, life cycle, and the functions of the various virally encoded proteins. According to my experimental results discussed in the followed chapters, many ISAV proteins express well in the bacterial expression system and can be easily purified. Therefore, ISAV can be used as an alternative model to study influenza viruses, especially for structural and functional studies of the viral proteins.
1.5 Anti-flu drugs targeting influenza proteins

Influenza is a major public health concern each year, and influenza A viruses can cause widespread pandemics with high mortality rates. The ultimate goal of studying influenza A virus is to treat the disease and to control the outbreak of epidemics. Therefore, drug design is always one of the most important topics in the field of researches on influenza viruses. By targeting the influenza proteins, many drugs can successfully control the propagation of the virus and treat the disease. As more and more structures of the influenza proteins are determined, we can expect more drugs to be developed based on the structures.

Developed by structure-based drug-design, the anti-flu drug oseltamivir (Kim et al., 1997) and zanamivir (von Itzstein et al., 1993) inhibit the influenza neuraminidase protein, by mimicking the sialic acid structure (Figure 1.12). The active ingredient of the anti-flu drug “TAMIFLU®” is oseltamivir phosphate. Oseltamivir and zanamivir were designed by chemically modifying sialic acid according to the binding pocket in the neuraminidase protein. These two drugs can effectively bind to the active site of the neuraminidase, causing it to be unable to release the nascent influenza virion from the host cell to infect...
Adamantane drugs amantadine and rimantadine have been found to target the influenza M2 protein as anti-viral drugs. However, the mechanism by which these drugs function remains unclear (Hu et al., 2010; Pinto and Lamb, 2006). Moreover, S31N, an adamantane-resistance mutation that greatly decreased inhibition by amantadine, has been developed on the influenza M2 proteins and now existed in more than 95% of circulating influenza viruses (Deyde et al., 2009; Krumbholz et al., 2009). New drugs against the influenza M2 protein (Wang et al., 2009) as well as the S31N mutants (Wang et al., 2013) have been reported, and they are likely to become new methods for treating influenza if their effectiveness can be confirmed.

As the most abundant viral protein, the influenza NP can also be a good target for anti-viral drugs. An appropriate oligomerization of the NP is critical for influenza virus replication, and thus the loop-binding pocket could a target for small-molecule drugs that mimic the structure of the tail loop to competitively inhibit oligomerization. On the other hand, drug compounds that promote aberrant NP aggregation can also effectively inhibit influenza A virus replication in cell cultures, also suggesting that the NP is a valid target for anti-influenza therapy (Gerritz et al., 2011; Kao et al., 2010).

Finally, the polymerase proteins can also be good targets for anti-flu drugs. A recent structural study showed that several known endonuclease inhibitors, including four
diketo compounds and a green tea catechin, bind to the endonuclease active site of the PA protein (Kowalinski et al., 2012). All of these inhibitors chelate the two critical manganese ions in the active site of the enzyme, although some differences are noted in the overall ligand orientation of these compounds. Further optimization of such endonuclease inhibitors may lead to potent drugs targeting the cap-snatching endonuclease activity of the influenza virus polymerase. Another promising approach to inhibit the replication of the influenza A virus is to disrupt the assembly of the viral heterotrimeric polymerase complex. It has been shown that short peptides derived from the N-termini of PB1 and PB2, which target the PA-PB1 and PB1-PB2 interaction interface respectively, exhibited varying levels of effectiveness in blocking the viral polymerase activity as well as the growth of the virus (Reuther et al., 2011; Wunderlich et al., 2009).
2. Methods

Structure determination of a target protein is often a long and challenging process that consists of multiple experimental steps. In the beginning, one usually needs to clone the target gene and inserts it into an expression vector. Next, expression needs to be tested and optimized to produce the protein of interest in an efficient, convenient, and economic way. Because crystallography requires very high protein purity, the expressed protein has to be purified from contaminating proteins from the expression host. After purification, a large-scale crystallization trial is usually setup with an automated system (e.g., crystallization robot) to screen over 1000 conditions using commercially available screening kits (e.g., Qiagen, Hampton research, Emerald Biosystems, and etc). Crystals from initial screens often need to be further optimized by varying a number of crystallization parameters, such as precipitant concentration, protein concentration, ionic strength, buffer pH, temperature, and so on. Seeding is usually a good technique to improve crystal quality. Sometimes, in order to further extend

Figure 2.1 Procedures to determine a protein structure from its gene.
the crystal diffraction limit, one needs to go through additional crystal manipulation steps, such as crystal annealing, crystal dehydration, and crystal cross-linking. Once the crystal diffraction is satisfactory, diffraction data sets are collected either from an in-house X-ray diffraction facility or a synchrotron X-ray source. To determine the structure, one needs the aid of several software packages such as HKL2000, Phenix, and CCP4. Finally the atomic structure is usually built with COOT. The entire process is summarized in the following flow chart (Figure 2.1).
2.1 Plasmid construction

The first step of plasmid construction is to decide which expression system to use. Usually this is a trial-and-error process, where one first starts with the prokaryotic (*E. coli*) expression system, and moves on to the insect cell system or the mammalian cell system if the *E. coli* expression system is not suitable. The *E. coli* expression system was first chosen for my study because of its simplicity, high yield, as well as its low cost. In addition, with the *E. coli* expression system, I could easily incorporate heavy atoms (e.g., Se by selenomethionine) into recombinant proteins, which would then allow phase determination using Multi-wavelength Anomalous Diffraction (MAD) (Hendrickson et al., 1990).

The gene of interest was cloned with pfu polymerase (Agilent Technologies, Inc.), with the corresponding cDNA as the amplification template. The PCR primers (synthesized by Sigma Genosys) used in my study introduced two restriction enzyme digestion sites, with NcoI or NdeI at the 5’ end and XhoI at the 3’ end. In the forward primer, a 6X His-tag was engineered by inserting the nucleotide sequence “5’-CACCACCACCACCACCAC-3’ ” onto the downstream end of the restriction enzyme digestion sites. A his-tag was chosen due to its small size, and was therefore expected to have only a minimal impact on protein crystallization. The PCR reaction (100uL) contained the following ingredients: 3.2 μL of both the forward and reverse primers (10 μM, Sigma Genosys), 2μL of dNTP (10 mM, Promega), 1 μL of cDNA template (5 ng/μL), 1 μL pfu polymerase (Agilent Technologies), 10 μL of 10X reaction
buffer (Agilent Technologies), and 79.6 μL of Milli-Q water. The PCR was performed in a 96-well Thermal cycler (Mastercycler® pro, Eppendorf) with 35 cycles of denaturation (95°C for 30 seconds), annealing (50°C for 30 seconds), and extension (72°C for 135 seconds), following by a final extension step at 72°C for 300 seconds. PCR products were analyzed by electrophoresis with a 1% agarose gel.

To prepare them for ligation, both the expression vector (pET28b+, Novagen) and the PCR products were digested with the appropriate restriction enzymes at 37°C for 4 hours. The digestion products were recovered by electrophoresis with a 1% agarose gel following by gel extraction with the QIAquick Gel Extraction Kit (Qiagen). The recovered pET28b+ backbone and the digested PCR product were mixed together at a molar ratio of 1:9 (plasmid to insert) for ligation with the T4 DNA ligase (NEB) at 16°C for 12 hours. The ligation product was transformed into DH5α competent cells using the heat-shock method, and cells were spread on LB agar plates with kanamycin (33μg/mL) and incubated at 37°C for 12 hours. Single colonies were picked from the LB agar plates and inoculated into LB broth medium with kanamycin (33μg/mL), and shaken at 250rpm at 37°C for 12 hours. A mini-prep was performed on the cell culture, and the plasmids obtained were analyzed by double digestion to confirm the size of the expression vector and the DNA inserted. After the sizes of pET28b+ backbone and the inserted DNA fragment were confirmed, the plasmid was sent for sequencing (SEQWRIGHT) to confirm the DNA sequence.
2.2 Protein expression

Recombinant plasmids with desired coding sequences were transformed into *E. coli*. Rosetta 2 (DE3) competent cells for protein expression. The Rosetta 2 (DE3) competent cell line was derived from the *E. coli* BL21 cell line. To enhance eukaryotic protein expression, tRNAs for seven rare codons (AGA, AGG, AUA, CUA, GGA, CCC, and CGG) were supplied by a chloramphenicol-resistant plasmid. The transformed cells were spread on a LB agar plate with kanamycin (33μg/mL) and chloramphenicol (30μg/mL) added, followed by incubation at 37°C for 12 hours. Single colonies were picked from the LB agar plate and inoculated into 50mL of LB broth medium with kanamycin (33μg/mL) and chloramphenicol (30μg/mL) added. The cell culture was shaken at 250rpm at 37°C until the OD$_{600}$ reached 0.7-0.8.

Isopropyl β-D-1-thiogalactopyranoside (IPTG) is generally used to induce protein expression (both native and SeMet-derivatives). IPTG is an analog of lactose that deactivates the lac repressor (LacI). LacI binds to the lac operator and inhibits gene expression. Naturally, when available, lactose is converted into allolactose to inhibit the LacI’s DNA binding ability. IPTG is a structural analog of the allolactose, and thus it can inhibit the inhibitor to allow protein expression. This is the so-called IPTG induction of protein expression. In addition, IPTG is not subjected to hydrolysis or host cell consumption, and therefore its concentration remains constant in the cell culture medium.
To express native proteins, once the OD\textsubscript{600} of the cell culture reaches 0.7-0.8, IPTG was added to the expression medium to a final concentration of 1mM to induce protein expression. The medium was usually shaken at 250rpm at a suitable temperature for up to 24 hours. Finally the cells were collected by centrifugation at 3,000g for 15min and stored at -80°C for expression analysis.

To express the SeMet protein for phase determination (discussed in the Phasing section, Chapter 2.6.2), the \textit{E. coli}. Rosetta 2 (DE3) transformants were inoculated in M9 minimum medium. When the OD\textsubscript{600} reach 0.8, 60 mg of SeMet (sigma) along with 100 mg each of threonine, lysine hydrochloride, and phenylalanine and 50 mg each of leucine, isoleucine, and valine were added as solids to the growing culture. After another 15 min, IPTG was added to 1 mM, and the procedure was completed as usual. Similar to native protein, SeMet expression was induced at 15°C for 24 hours before cell collection similar to native proteins.

To extract the recombinant protein, the harvested cells first need to be lysed. Sonication is a popular technique for this purpose where cells are lysed by liquid shear and cavitation. In my experiments, a Branson Sonifier 250 was used for sonication. Depending on the volume of the cell lysate, different durations of sonication were needed to ensure complete cell lysis. Usually, 3 x 1 min cycles of sonication were sufficient for a 50mL cell culture, whereas 3 x 10 min cycles of sonication were needed for cell cultures that were greater than 1L in volume.
To analyze the protein expression, the cell pellet of the 50 mL culture was resuspended on ice with 5 mL of lysis buffer containing 50mM Tris pH7.5, 300mM NaCl, 10% glycerol (v/v), 5mM imidazole, 17μg/mL phenylmethylsulfonyl fluoride (PMSF, a serine protease inhibitor), and 5 mM β-mercaptoethanol (2-ME). 3 x 1 min cycles of sonication were performed on ice with a Branson Sonifier 250. A 20μL whole-cell lysate sample was taken for analysis later. The sonicated sample was centrifuged at 22,000g for 30min at 4°C to separate the supernatant from the pellet. A SDS-PAGE was performed to analyze the protein content of the whole-cell lysate, supernatant (soluble protein), and pellet (insoluble protein). If the expression level and the solubility of the recombinant protein were good, a large scale protein expression experiment (usually 4L-8L) would be performed to produce more protein for purification.

2.3 Protein purification

To obtain high quality crystals for structure determination by X-ray crystallography, it is necessary to obtain a large amount of protein with high purity. Thanks to the advances in chromatography media and instrumentation, the efficiency and ease of protein purification have greatly improved during the past decades. Nowadays we have a variety of purification resins, pre-packed columns, and different instruments to choose from. In our lab, we use the "ÄKTA FPLC" system from GE Healthcare Life Sciences (Figure
2.2A). The common steps taken to purify a target protein usually involve affinity purification using Ni-NTA and heparin resins, ion-exchange purification using pre-packed column (Figure 2.2 B), and finally size-exclusion chromatography based on molecular size (Figure 2.2 C).

An N- or C-terminal His-tag attached to the target protein makes affinity purification possible with Ni-NTA resin. Histidine residues in the His tag bind to the immobilized nickel ions on the resin with high specificity and affinity. This is the step where most of the contaminating proteins can be removed because only proteins with exposed multi-histidine sequences are able to bind to the Ni-NTA resin with a strong affinity.

Figure 2.2 FPLC instrument and columns for chromatography. A. ÄKTA FPLC system from GE Healthcare Life Sciences. B. Different kinds of pre-packed columns for affinity chromatography. C. Superdex 200 gel-filtration columns for size-exclusion chromatography. All figures from www.gelifesciences.com.
To purify the His-tagged protein, the supernatant of the cell lysate was incubated with Ni-NTA resin for 1 hour at 4°C. The mixture was gently shaken to facilitate the protein binding to the resin. After incubation, the mixture was centrifuged at 300g for 5min at 4°C. The resin was resuspended with washing buffer (50mM Tris pH7.5, 300mM NaCl, 10% glycerol (v/v), 20mM imidazole, 17μg/mL PMSF, and 5mM β-mercaptoethanol) and loaded into a gravity column for elution. Once the resin was thoroughly washed, elution buffer (50mM Tris pH7.5, 300mM NaCl, 10% glycerol (v/v), 1M imidazole, 17μg/mL PMSF, and 5mM β-mercaptoethanol) was added to the column and the eluted protein was collected in fractions. A SDS-PAGE analysis was done in the end to check the protein contents of the whole-cell lysate, supernatant, pellet, flow through, wash, and eluted fractions.

The next step was to purify the protein with affinity chromatography with pre-packed columns (GE Healthcare Life Sciences). The column could be ion-exchange columns, either cation- or anion-, depending on the net surface charge of the target protein. Also, a Heparin column (GE Healthcare Life Sciences) was often used to purify proteins capable of binding to nucleic acids, because the Heparin resin was a polymer of sulfated disaccharide, whose chemical properties resembled that of nucleic acids. For the ISAV-NP project, the Heparin column was used right after the Ni-NTA resin.

To purify the protein with pre-packed affinity columns, the eluted fractions with the target protein from the Ni-NTA resin were pooled together and dialyzed into a buffer
containing as little salt as possible. The dialyzed protein solution was injected into the pre-packed affinity column with a flow rate of 2 ml/min. The protein bound to the column was eluted using the ÄKTA FPLC system (GE Healthcare Life Sciences), with a concentration gradient of sodium chloride varying from 0 M to 1 M. All eluted fractions with protein were analyzed by SDS-PAGE to determine in which fractions the target protein was eluted.

In the end, size-exclusion chromatography was performed. This step did provide much help in removing contaminating proteins. Rather, it was most helpful for separating different oligomeric species of the target protein, thus improving sample homogeneity for crystallization. Also this step could serve as a good quality control step for the whole purification process, since it could be used to determine if the protein was of the correct size, whether the oligomerization was consistent with expectations, and whether the protein was properly folded, based on the elution volumes and the shapes of the peaks in the size-exclusion chromatography chromatogram.

The size-exclusion column used was Superdex 200 16/60 (GE Healthcare Life Sciences). The column was pre-equilibrated with gel-filtration buffer containing 25mM Tris pH7.5, 200mM NaCl, 10% glycerol (v/v), and 5mM β-mercaptoethanol. The protein purified from the previous step was concentrated to 2mL and injected into the injection loop. The column was eluted with the gel-filtration buffer, and the eluted fractions were analyzed by SDS-PAGE. If pure protein was obtained with a symmetric, well-defined
peak, the protein was concentrated to 10 mg/mL for the subsequent screening of crystallization conditions.

2.4 Crystallization

Once a recombinant protein has been purified to homogeneity, the next rate-determining step in obtaining a crystal structure is the crystallization of the protein. Although there are many crystallization methods available (Bergfors, 1999; Bergfors, 2009; Ducruix and Giege, 1999; McPherson, 1999), vapor diffusion is the most commonly used one and it has led to more proteins being crystallized and structures being solved than all of the other methods combined (McPherson, 1999).

Vapor diffusion can be set up in two ways, the hanging-drop and the sitting-drop (Figure 2.3 A). In both set-ups, the drop containing the protein solution and the mother liquid is equilibrated against a reservoir (mother liquid) such that the ingredients of the protein drop are gradually concentrated until evaporation equilibrium between the drop and the well solution is reached. In Figure 2.3 B, point “a” corresponds to an initial setup state. At this point the concentration of protein is too low for crystallization (undersaturation), so the drop is clear. Because the precipitant concentration in the drop is lower than the reservoir, water diffuses from the drop so the concentrations of both the protein and the precipitant increase in the drop (moving from point “a” to “b”). Once the protein concentration reaches point “b”, it is in the nucleation zone and spontaneous nucleation
occurs. As the protein goes from liquid phase (solution) to solid phase (crystal), the protein concentration decreases (moving from “b” to “c”). During this process, more and more protein molecules pack on the crystal surface, so the sizes of crystals increase and they become visible.

**Figure 2.3** Protein crystallization. A. Hanging-drop and sitting-drop set-up of the vapor diffusion method. B. A schematic drawing of a protein crystallization phase diagram based on the protein concentration and the precipitant concentrations.
Vapor diffusion has been the crystallization method used in my thesis project. Both hanging-drop and sitting-drop settings have been used at various occasions. Obtaining high quality protein crystals involves two stages: large-scale screening and manual optimization. The large-scale screen was performed with the “Hydra II Plus One” robot, which was able to screen over 1000 crystallization conditions that were commercially available in the 96-well tray format (Figure 2.4 A-C). The commercial crystallization screens were purchase from Hampton Research (including Index, Crystal Screen, PEG/ION, and SaltX), Qiagen (including MbClass I/II, pHClear I/II, Classic, and JCSG+), and Emerald Biosystems (including Wizard I-IV and Cryo I/II). After possible crystallization conditions (“hits”) were obtained, 24-well trays (Hampton Research) were

Figure 2.4 A. Example of a crystallization robot, controlled by a computer. B. Commercial crystallization screen containing 96 different conditions. C. A 96-well tray for the initial large-scale screens. D. A 24-well tray for the optimization of the crystallization condition. Figure A taken from acenewsevents.blogspot.com. Figure B taken from www.omscientia.com. Figure C, D taken from hamptonresearch.com.
used to manually and systematically optimize crystallization conditions (Figure 2.4 D) to obtain quality crystals for X-ray diffraction. To optimize the condition, conditions were designed to vary precipitant concentration, ion strength, buffer pH, temperature, and protein concentration. Additives and detergents were sometimes included to improve the crystallization process and/or the quality of the crystal diffraction. Additional post-crystallization treatments also helped to increase the diffraction limit of the crystals (described later).

X-ray diffraction is often performed at cryogenic temperature (~100 K) to reduce the radiation damage to the crystal during the diffraction (Alcorn and Juers, 2010). Protein crystals usually contain around 50% water. Therefore, they need to be “cryo-protected” before performing the diffraction experiments. If the cooling is not fast enough, the water in the protein crystal can turn into ice crystals, and the large volume change during the solidification to ice may damage the crystal lattice and negatively impact the diffraction. However, this can be avoided if the crystals are cryo-protected prior to diffraction and then subjected to fast cooling (Kriminski et al., 2002). When cooling is performed under cryo-protected conditions, water in the protein crystal form amorphous ice, during which process the volume change is very little such that the crystal lattice is not damaged. Choosing a good cryo-protectant is critical for good diffraction and the choice varies from case to case. For my experiments, I supplemented 20% (v/v) glycerol or PEG 400 to the mother liquor recipe to make it “cryo”. This method has worked for all of my crystals so far.
2.5 Crystal manipulation

Unless one gets extremely lucky, the crystallization conditions obtained from the initial screen need to be optimized in order to get satisfactory X-ray diffraction. To optimize a condition, one will usually vary the protein concentration, precipitation concentration, buffer pH, salt type/concentration, temperature, and other parameters. Additive screens and detergent screens can also be performed to see if the protein can be better crystallized in the presence of a certain additives/detergents. However, it is often the case that further improvement is still needed even after all of the methods previously mentioned have been implemented. Here, I describe four useful crystal manipulation techniques that may further enhance the quality of the X-ray diffraction: crystal annealing, crystal dehydration, crystal crossing linking, and seeding. All of these are useful tools when crystals can be obtained but unfortunately produce low quality diffraction, and thus these methods are also called post-crystallization treatments. In fact, among them, crystal dehydration and seeding worked well in this thesis project and improved the diffraction to 2.7Å which led to successful structure determination.

2.5.1 Crystal annealing

Although fast-cooling of the cryo-protected crystals is critical, it can disrupt the crystal lattice which leads to higher mosaicity and lower diffraction resolution (Rodgers, 1994). This problem is believed to be due to uneven cooling which results in different expansion rates of the crystal versus the solvent. This introduces tension in the crystal which distorts the crystal lattice to some extent. It is more often seen with large crystals...
mounted in a large loop. Crystal annealing may solve this problem. It is a simple one-step method which treats the crystal in-place: first the cold stream is blocked to let the crystal return to room temperature, and then flash cooled again (Harp et al., 1998). This approach may restore diffraction to flash-cooled crystals that were not so well frozen previously.

### 2.5.2 Crystal dehydration

The crystal dehydration is a very powerful method to enhance crystal diffraction by inducing structural changes and potentially improving crystal packing. Dehydration removes excess solvent and tightens the packing of protein molecules which may make the packing more ordered. Therefore crystal dehydration could be especially useful for crystals with high solvent contents. There have been lots of examples reported about diffraction improved by crystal dehydration (Heras and Martin, 2005).

![Figure 2.5](image_url) **Figure 2.5** Schematic drawing of crystal dehydration process by serial transfer of the cover slide. Modified from Heras et al (Heras and Martin, 2005).

I used a “serial-transfer” method to do crystal dehydration as shown below (Figure 2.5). After the crystal grew to full size, I transferred the cover slide to a reservoir containing mother liquid with a higher concentration of precipitant, and then incubated it for 8-12
hours to let the drop and mother liquid equilibrate. Further steps of transfer and equilibration were performed if necessary. Also, the concentration increments were adjusted accordingly. For example, for a more gentle dehydration, the increment was reduced to 1-2% for each step.

2.5.3 Crystal cross-linking

Crystal cross-linking is a gentle method which introduces glutaraldehyde into the drop containing the crystals by vapor diffusion. Cross-linking can increase the strength of the internal structure of crystals such that the damage to the crystal lattice during fast-cooling may be reduced. For my cross-linking experiments, a micro-bridge (Hampton Research) was used to hold 5μL of 25% glutaraldehyde, and the micro-bridge was put in the well for crystallization (Figure 2.6), glutaraldehyde can accumulate in the drop via vapor diffusion and react with the proteins molecules of the crystal. The equilibration took 60 minutes before the crystals were tested with X-ray diffraction experiments.

Figure 2.6 Schematic drawing of crystal cross-linking process by serial transfer of the cover slide. Modified from Heras et al. (Heras and Martin, 2005).
2.5.4 Seeding

As shown in the previous phase diagram (Figure 2.3), the process of protein crystallization can be divided into two steps. The first step is nucleation, where tiny nuclei of “micro crystals” spontaneously form when the concentration of protein and precipitant are sufficiently high. Once the some protein goes from liquid phase (solution) to the solid phase (nuclei), the protein concentration in solution goes down. At some point, spontaneously nucleation is no long possible, and the process goes into the second phase -- crystal growth -- where the protein molecules keep packing on the surfaces of the existing nuclei, but the number of crystal stays constant since then. The concentration of protein required to pack on to the existing crystal is lower than that the amount of protein needed to spontaneously form nuclei, because the latter is a kinetically demanding step.

Spontaneous nucleation is a step that is hard to control. It is no good if there is no nucleation, or if there is too much nuclei in a drop. The former leads to no crystallization, and the latter causes a crystal shower -- where too many nuclei cause the protein to get depleted too fast and the sizes of the resulting crystals are too small. Seeding solves this problem. At a protein concentration where crystal growth is possible but spontaneously nucleation is not, a few nuclei, the number of which is “controlled”, are introduced to the drop. Protein will only accumulate on the surface of the exogenous nuclei and the resulting crystals will be large in size. Because one can control the rate of crystal growth by adjusting the initial concentration of protein/precipitant, the internal packing of the
crystals may also be improved by seeding. Therefore, with seeding, one may obtain fewer, bigger, and better packed crystals.

When I performed the seeding experiments, protein crystals were first crushed with seeding beads (Hampton research) using the vortex. The seed was then diluted from 10 times to $10^4$ times, and introduced to pre-equilibrated drops whose protein concentration and/or precipitant concentration were lower than that required for spontaneous crystallization. Multiple parallel crystallization experiments with such gradients were usually set up to determine the best seeding condition.
2.6 X-ray crystallography

X-ray crystallography is the most commonly used method to determine the atomic structure of biological macromolecules. X-ray crystallography is essentially a form of very high resolution microscopy. Although we cannot directly “see” the protein structures, we can re-construct a structural model of the proteins molecules from the X-ray diffraction pattern using Fourier transforms. In all forms of microscopy, the resolution is limited by the wavelength of the electro-magnetic radiation used, and this is why we use X-rays (wavelength ~1Å) to obtain structural information at the atomic level. Although even single molecules can diffract X-rays, the diffraction is too weak to be distinguished from noise. Crystals, an ordered three-dimensional array of molecules, can magnify the signal to make it detectable. That being said, it is critical to have a well-ordered crystal for X-ray crystallography, in order to diffract the X-ray to high angles which is necessary to reveal detailed information about the protein structure.

2.6.1 Data collection

During an X-ray diffraction experiment, the protein crystal, which is exposed to the beam, diffracts the X-rays and a diffraction pattern can be collected by the CCD.

![Figure 2.7](image.png)

**Figure 2.7** Schematic drawing of X-ray beams being reflected by the atoms in the crystal. Constructive interference occurs when $2d_{hkl} \sin \theta = n \lambda$ (Rhodes, 2006).
detector. We see “spots” on the diffraction images, and those spots correspond to constructive interference of the electromagnetic waves of the diffracted X-rays. The crystals can be viewed as sets of parallel and equivalent planes that are able to reflect X-rays (Figure 2.7). According to Bragg’s law, in an X-ray diffraction experiment, constructive interference occurs if and only if the following equation is satisfied:

\[ 2d_{hkl} \sin \theta = n\lambda \]

Where \( d_{hkl} \) is the inter-plane distance, \( \lambda \) is the wave length of the X-ray beam, \( \theta \) is the reflection angle, and \( n \) is an integer.

![diagram](http://hazeslab.med.ualberta.ca/)

**Figure 2.8** Schematic drawing to show that diffraction occurs when and only when a reciprocal lattice point intersects the Ewald's sphere. Figure taken from http://hazeslab.med.ualberta.ca/

According to Bragg’s law, all of the X-ray reflections can be reconstructed with arrays of
lattice points in reciprocal space. During data collection, the crystal is rotated about its z-axis, which makes its reciprocal lattice points intersect with the sphere of reflections, the Ewald’s sphere with a radius of 1/\( \lambda \). A diffracted beam would be produced once the reciprocal lattice points intersect the Ewald’s sphere, as shown in the figure above (Figure 2.8). During data collection, the crystal is periodically oscillated a small amount (e.g., 0.5° or 1°). Since the reciprocal lattice rotates exactly the same as the rotation of the crystal, the crystal’s oscillation ensures that the Ewald’s sphere intersects with as many of the reciprocal lattice points as possible. Therefore, the oscillation of the crystal maximizes the unique reflections that could be collected. All reflections can be recorded by crystal rotation and oscillation except those blind regions which are caused by the curvature of the Ewald’s sphere (also called the cusp). In theory, blind regions exist for any data collection, but these blind regions may be related by symmetry to the volume of reciprocal space that gets measured, and thus the data set is complete and no cusp data collection is necessary in these cases.

In most cases, a complete sphere of data can be collected on any crystal by rotating the crystal through 180 degrees. However, 180-degree collection is not always required, and the required minimum oscillation range is related to the space group symmetry of the crystal lattice, as well as the crystal orientation. There are 14 different Bravais lattices (Figure 2.9) and 32 point groups, which combine together gives 230 different groups. Among those, only 65 space groups without mirror planes and inversion centers are possible for protein crystals. The triclinic space group, the space group with the lowest
symmetry, requires 180-degrees to be collected for a complete data set. Spaces group with higher degrees of symmetry, for example, the hexagonal space group P622, may need only as little as 30-degrees to be collected in order to be complete. If the crystal is robust enough and time permits, it is always desirable to collect a 180-degree data set. However, if the radiation damage is a problem or there is time restraint, one should consider a data collection strategy to maximize the resolution while maintaining the completeness of the data.

Prior to the collection of a full data set, it is critical to determine the conditions to most efficiently collect the data. There are at least four factors to decide: distance, exposure
time, oscillation angle, and oscillation range.

a. **Distance.**

A longer distance between the crystal and the image plate is beneficial in that it reduces the amount of overlapping spots, whereas a shorter distance is favored if high-resolution data needs to be captured. Usually, it is the best practice to adjust the distance such that the edge of the diffraction pattern almost reaches the edge of the detector, which strikes a balance between capturing high-resolution data and avoiding overlapping. Another consideration is that the distance cannot be so far that the absorption of X-rays by the air is significant.

For my thesis project, a 170mm distance was chosen based on the resolution (3Å) determined from the snapshot, and the separation distance between the spots shown on the snapshot images.

b. **Exposure time.**

To determine the amount of time to expose the crystal to the X-ray beam, one need to consider another trade off: intensity and radiation damage. If the intensity is higher, it is possible to see some spots at higher resolutions that would be otherwise invisible. Therefore, a longer exposure time may lead to a final model with higher resolution. However, on the other hand, the risk of longer exposure is that the radiation damage may deteriorate the diffraction ability of the crystal over time, such that the crystal may not even survive till the end of the entire data collection. This
may lead to a low completeness of data, or even failure in determining the structure. Since one can never guess how long a crystal can survive in an X-ray diffraction test, it is desirable to start with some not-so-good crystals to get a rough estimation.

c. Oscillation angle

As discussed before, during the collection of each single image, the crystal is rotated back and forth (oscillated) to allow the reciprocal lattice points to intersect the Ewald’s sphere so that the number of the collected unique reflections is maximized. The oscillation angle has a positive relationship with the number of reflections (“spots”) collected on an image. To determine how large the oscillation angle should be, one needs to take a snapshot of the crystal and estimate the unit cell dimension as well as the mosaicity. There are more reflections detected for crystals with large unit cell dimensions than those with small unit cell dimensions. Thus, it is usually beneficial to reduce the oscillation angle, if the unit cell dimensions of the crystal are large, to avoid overlapping the reflections. Mosaicity is a metric used to measure the degree of long-range order of the unit cells within a crystal. Lower mosaicity indicates better ordered crystal packing, whereas high mosaicity broadens the reflection spots such that the chance of overlapping is higher. For protein crystals with relatively high mosaicities (above 1.0), a small oscillation angle (e.g., 0.5 degree) needs to be chosen. However, if the oscillation angle is too small, the crystal may not survive long enough for sufficient frames to be collected in order to obtain adequate data completeness. Therefore, one should consider the crystal mosaicity, the potential
radiation damage to the crystal, and the space group of the crystal before making a careful decision.

d. Oscillation range

The oscillation range is how many degrees the crystal is rotated in total during the data collection. Obviously, the oscillation range is a critical factor for determining the data completeness and redundancy, and it is also related to resolution. There is nothing wrong with collecting as large an oscillation range as possible, but one may rationally revise this strategy in order to minimize radiation damage to crystals. It is usually a good practice to first take a snapshot of the crystal in order to determine its space group, and then rationally design a good data collection strategy, instead of collecting 180-degrees of data on every crystal with the same strategy. Although it does no harm to collect 180-degrees of data for robust crystals that are resistant to radiation damage, one may consider decreasing the oscillation range and increasing the exposure time per frame for small crystals that have higher symmetries but are sensitive to radiation damage. In this way, high resolution datasets can be collected without affecting the completeness of the data set.

As mentioned above, generally all my crystals had a relatively high mosaicity such that the oscillation angle had to be small, e.g., 0.5°, to avoid overlapping. On the other hand, it seemed that my crystals were also prone to radiation damage, so collecting 180 degrees of data with 0.5 degree per frame was not possible. After indexing the initial snapshot
image, the crystal space group was determined to likely be P422 which has a relatively high symmetry, and therefore obtaining data from a large oscillation range was not necessary. In the end, my data was collected with a 170mm distance, an oscillation angle of 0.5 degree, an oscillation range of 75 degree, and an exposure time of 4 seconds per frame.

2.6.2 Phasing

One of the most important steps of X-ray crystallography is the generation of an electron density map, with which the atomic model of the protein can be built. The calculation of the electron density follows this equation:

$$\rho(x, y, z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F_{hkl}| e^{i(2\pi(hx + ky + lz) - \alpha)}$$

where $V$ is the volume of the unit cell; $h$, $k$, $l$ are the Miller indices of a reflection $hkl$, $|F_{hkl}|$ is the amplitude and $\alpha$ is the phase angle. As shown, the indices of a reflection $(h, k, l)$, the intensity of the reflections, $|F_{hkl}|$, and the phase angles of the reflections, $\alpha$, are needed to calculate the electron densities from a diffraction experiment. However, only two (indices and intensities) of which can be determined directly by the experiment, and the phase information is missing. This is the so-called phase problem of X-ray crystallography.

There are four major methods to solve this phase problem:

a. Direct methods

This method has pretty limited usage in crystallography for macromolecules, but it is
still worth mentioning. By using the direct method for phasing, one assumes that a crystal is made up of atoms of similar shape as well as with positive electron density, and thus there are statistical relationships between sets of structure factors. Direct methods look at such relationships to deduce possible values for the phases. Direct methods are the most useful with data sets that have a small unit cell with only a few hundred atoms. However, it is increasingly more difficult for this method to be applied with a larger unit cell. Moreover, direct methods require extremely high resolution (at least 1.2Å), which is generally hard to reach for protein crystals.

b. Molecular replacement (MR)

If the protein of interest (or part of it, e.g., a domain) has a homolog with a known structure, it is usually a good idea to start phasing with molecular replacement. Usually it is straightforward to carry out molecular replacement is there is 40% or above sequence homology between the protein and the search model, but this does not mean that molecular replacement is impossible if there is no sequence identity. During molecular replacement, the search model (coordinates of the known structure) is placed in the unit cell to search for the correct orientation and position. Usually the rotational function is first computed to determine the three rotation angles, and then the oriented model can be placed in the unit cell to compute a 3D translation function for the three translation parameters (position in the unit cell). After the search model is correctly positioned inside of the unit cell of unknown structure, the initial phases can be estimated.
c. **Multiple isomorphous replacement (MIR)**

By soaking, multiple isomorphous replacement introduces heavy atoms (gold, platinum, mercury, and etc.) to the crystal and the heavy atoms change the scattered intensity significantly because they contribute disproportionately to the overall intensity due to their much bigger sizes compared to the common atoms of a protein. For successful phasing with heavy atoms, the crystal derivatives with heavy atoms should be still comparable to the native crystal, meaning that they share the same space group, similar unit cell dimensions, and similar resolutions. The comparison of the scattering intensities of the native crystal and the heavy-atom derivative crystal allows the computation of isomorphous difference Patterson map and determination of the heavy atom positions. With this information, one can compute the contribution of the heavy atoms to the structure factors, and estimate the possible values of the protein phase angles.

d. **Anomalous dispersion**

In most cases, the reciprocal lattices have a center of symmetry, where the intensity of \((h k l)\) and \((-h -k -l)\) (Friedel’s pair) are equal according to Friedel’s law. However, when the wavelength of the X-ray is at the edge of an element’s absorption, or the wavelength of the X-ray corresponds to a transition between different electron shells of the atom, there will be anomalous scattering and a modification of the phase. This causes the diffraction pattern to be distinguishable from that of the native crystal. The positions of the atoms producing the anomalous scattering can be calculated,
and the initial phase angles can be estimated with the location of the atoms giving the anomalous scattering.

For an anomalous dispersion experiment, selenomethionine (SeMet)-derivative crystals are most commonly prepared and the diffraction experiment is performed with the X-ray wavelength that is at the edge of the absorption of the selenium atom. To express SeMet protein, auxotrophic met- strains were previously used with a supplement of exogenous selenomethionine supplied. Recently, a technique for inhibiting the methionine biosynthesis pathway has been developed. This method can be used with any expression vector and is applicable to any prokaryotic strain. The methionine biosynthesis can be blocked by inhibiting aspartokinases, in the presence of high concentration of isoleucine, lysine and threonine (Benson et al., 1995; Doublie et al., 1996; Ducruix and Giegé, 1999; Hendrickson et al., 1990; McPherson, 1999; Van Duyn et al., 1993). The detailed procedures for expression of SeMet-protein have been discussed in Chapter 2.2, protein expression.
2.6.3 Structure determination

To process the data, the reflection data from the images first need to be indexed, integrated, and scaled to obtain the three dimensional reciprocal space indices (h, k, l), and their corresponding diffraction intensity ($|F_{hkl}|$) with the estimated error ($\sigma$). HKL2000 was used for this purpose. The HKL2000 suite is a package of programs intended for the analysis of X-ray diffraction data collected from single crystals, and consists of three programs: XdisplayF for the visualization of the diffraction images/patterns, Denzo for parameter refinement, data reduction, and integration, and Scalepack for merging and scaling the intensity information obtained by Denzo, and generating the final statistics (Otwinowski and Minor, 1997).

First, the X-ray source was specified such that the beamline-specific parameters can be automatically read in (e.g., beam center). The raw data directory was specified to allow the program to read in all the images (.img format), together with the data collection parameters (oscillation angle, exposure time, distance, and etc.) which were stored in the image files. A path to the output directory was also given to the program for the storage of the results after scaling. A random reflection image was chosen for the initial indexing. To start the indexing, a peak search was performed to instruct the program to locate all the reflections. Peak searches could be automatic or manual, but usually they were done automatically. The values of the box sizes and spot sizes needed to be adjusted carefully, with the aid of a zoom window, to avoid overlapping and maximize the signal-to-noise ratio as much as possible. After peak searching, clicking the index button generated a list
of Bravais lattices. Usually, the Bravais lattice with the highest symmetry and lower than 1% distortion (highlighted in green) was chosen for subsequent processing.

Once the Bravais lattice was decided, refinement was performed. There were many refinement options listed on the left panel of the GUI, and I selected only the “Fit Basic” for the first-round refinement. Refinement was repeated until the number of the errors (the number in the middle column of the bottom panel in refinement information) converged to zero. Similarly, a second-round refinement was performed with all refinement options except “RotZ” and “Mosaicity”, and the third-round refinement was performed to include the “Mosaicity” options. The “RotZ” was never selected because the crystal did not rotate about the Z-axis. This step-wise refinement made it much easier to get a result with all error terms converged to zero. If one tried to refine all options at one time, sometimes the error would never converge.

During the refinement, it was important to monitor if the predicted reflections agree with the actual ones on the images (agreement shown in cyan for full reflection, yellow for partial reflection, and red for discarded reflection). Also, the $\chi^2$ values (positional and partiality) needs to be less than 1 (green indicates good values, orange indicates problematic, and red indicates disallowed).

Next, integration was performed to integrate all the remaining images with the refined parameters, and the reduced reflections from all images were scaled and merged together.
Because the SeMet protein was used to collect anomalous data, the “Anomalous” option was checked. The “Write rejection file” option was also checked to record the rejected reflections for the subsequent scaling operations. Since the second run of scaling, the “Use rejections on next run” was check to include the rejection information from previous runs. Scaling was repeated until the percentage of reflections marked for rejections equals the percentage of reflection rejected. During scaling, the error models might be adjusted to keep the normalized $\chi^2$ values around 1. After scaling, a .sca file was generated which contained all of the structural data from the X-ray diffraction. The .log file was examined in order to understand the quality of the data set. The following four criteria should be checked:

a. $R_{\text{merge}}$

$R_{\text{merge}}$ was used to determine the resolution of the data set. The resolution was “cut off” if the $R_{\text{merge}}$ value exceeds 50%. For a good data set, the average $R_{\text{merge}}$ value for all resolution bins should be less than 10%, and the $R_{\text{merge}}$ value for the highest resolution shell should be no more than 50%.

b. $I/\sigma$ (signal to noise ratio).

This was another, yet less commonly used, metric to determine the resolution of the data set. The resolution should be “cut off” before the $I/\sigma$ value goes to lower than 2.0. The higher this ratio was, the more confident we were on the validity of the reflections.
c. Completeness

This metric measured the percentage of the possible unique reflections that have been collected. 90% was required for protein crystals. Datasets that were less than 85% complete were rarely useful for anything.

d. Redundancy

Redundancy showed the average number of symmetry-related observations for each unique reflection that would be recorded, assuming that anomalous data is not needed. The redundancy of anomalous measurements would be lower. As the redundancy went up, the final averaged data quality definitely improved, but the $R_{merge}$ also went up.

After a good data file (.sca) was generated from HKL2000, Phenix was used for structure determination and model building. Phenix was an UNIX-based, highly automated system for macromolecular structure determination (Adams et al., 2010). Its functionality included maximum-likelihood molecular replacement (PHASER), heavy-atom search (HySS), template and pattern-based automated model-building (RESOLVE), automated macromolecular refinement (phenix.refine), and iterative model-building, density modification and refinement that could operate at moderate resolution (RESOLVE, AutoBuild). There was also a fully automated pipeline (Autosol) to integrate all of the necessary steps between supplying the data file and obtaining the final refined model, to ease the process of structure determination. Autosol has been the primary program used for the structural determination of the ISAV-NP.
Autosol was an experimental phasing pipeline that combines HySS (Hybrid Substructure Search) for finding heavy-atom sites, Phaser or SOLVE for calculating experimental phases, and RESOLVE for density modification and model-building. It could automatically try multiple solutions as necessary and determines which is most likely to give the desired result. As a first step, Autosol ran phenix.xtriage, a data analysis program, to identify potential problems such as twinning, translational pseudo-symmetry, or weak anomalous signal. A warning would be raised if any problems were detected by the Autosol. The second step of Autosol was to do a heavy-atom search, with the HySS program. When phasing with SAD or MAD, there could be multiple solutions obtained, and Autosol will proceed with the most probable one. The third step, phasing with SAD, was done with the Phaser program. Once the phases were obtained, the electron densities were be calculated by the density modification program RESOLVE. If the “autobuild” option was checked, the RESOLVE program would try building amino acid residues into the electron density, with the help of the user-supplied sequence file. When phasing with Se-SAD, the SeMet residues were also built according to the heavy-atom positions. After auto-building, COOT was used to manually add residues that were not built, and modify residues that were built wrong. Refinement with phenix.refine was done to generate the $R_{\text{work}}$ and $R_{\text{free}}$ values to evaluate the quality of the final model. In the beginning of the refinement step, about 10% of the experimental data (randomly chosen) was removed from the data set, and refinement was performed with the remaining data (90%) which yielded the $R_{\text{work}}$ value. Then, the $R_{\text{free}}$ value was calculated by measuring how well the model could predict the 10% preserved data that was not used in
refinement. Ideally, the $R_{\text{work}}$ value and $R_{\text{free}}$ value should be close to each other, if the model did not over-interpret the data.
2.7 Biophysical experiments

To better understand the interaction between the ISAV-NP and RNAs, biophysical tools were employed. In this thesis project, fluorescence polarization assays (FP) was used to determine the affinity and stoichiometry of the NP RNA binding with synthetic RNA oligos. Electron microscopy (EM) and dynamic light scattering (DLS) experiments were employed to measure the size of NP-RNA complexes as indirect approaches to confirm the binding stoichiometries.

2.7.1 Fluorescence polarization

Fluorescence polarization (FP), which is also called Fluorescence Anisotropy (FA), is used to characterize the interaction between protein and nucleic acids. Usually the nucleic acids are labeled with fluorophores, and the protein is unlabeled. Each fluorophore has an excitation dipole and an emission dipole. If the fluorophores are treated with

![Diagram of fluorescence polarization](image)

**Figure 2.10** Schematic presentation of how protein decreases the anisotropy of the emitted fluorescence from the fluorophore-labeled nucleic acids (LiCata and Wowor, 2008).
polarized light, only a portion of the fluorophores, whose excitation dipoles align with the incident polarized light, would be excited. If there is no delay between the absorbance of the exciting photon and fluorescence emission, the emitted fluorescence would also be completely polarized. But in fact, there is a delay between the excitation and emission, during which time the molecules tumble resulting in the emitted fluorescence being less polarized. The smaller the molecule is, the faster the molecule tumbles, and the less polarized the emitted light would be. In a protein-nucleic acid setting, the nucleic acids are labeled with fluorophores. Free nucleic acids are small and they tumble and rotate fast, so the emitted fluorescence is less polarized. When nucleic acids are bound by protein this slows the tumbling and rotation making the emitted fluorescence more polarized. The tumbling and rotational motion of fluorophore-tagged nucleic acids is slowed down by protein binding, resulting in higher levels of polarized fluorescence emission from the bound nucleic acids. Therefore the amount of nucleic acids that are bound to the protein can be quantified, and the protein-nucleic acid binding affinity can then be calculated.

5'-fluorescein-labeled RNA oligos were purchased from Thermo Fisher Scientific and Sigma-Aldrich. The sequences of the RNA oligos used are simple “AC” repeats (e.g., a 12mer RNA has the sequence of 5’–(AC)_{12}–3’) to avoid potential secondary structures that may interfere with protein binding.

For the affinity measurements, our binding solutions contained 100mM NaCl, 25 mM
Tris, pH 7.5, 10% glycerol, and fluoresceinated RNA at 0.2 nM or higher (concentrations were kept at least 10-fold lower than the $K_d$). NPs were titrated into the binding solution until the millipolarization (mP) stabilized. All of the experiments were carried out at 25°C. The data were plotted and analyzed by using the equation $P = (P_{\text{bound}} - P_{\text{free}})[\text{protein}]/(K_d + [\text{protein}]) + P_{\text{free}}$, where $P$ is the polarization measured at a given total protein concentration, $P_{\text{free}}$ is the initial polarization of fluorescein-labeled RNA without protein bound, $P_{\text{bound}}$ is the maximum polarization of RNA when all are bound by protein, and [protein] is the protein concentration. The free and total protein concentrations were assumed to be equal since the $K_d$ was at least 10-fold higher than the concentration of fluorescein-labeled RNA. The hyperbolic curves were fitted by nonlinear least-squares regression analysis assuming a bimolecular model such that the $K_d$ values represent the protein concentrations at which half of the ligands are in the bound-state.

For the FA experiments used to determine the binding stoichiometry, the assay conditions were identical to those used for the binding affinity determination, except that the RNA concentration was increased to 60 nM and the NaCl concentration was reduced to 50mM. These adjustments were made to ensure that the RNA concentration was at least 20-fold higher than the $K_d$. The titration curves showed a linear increase (first phase) in millipolarization until the RNA-binding sites were saturated (second phase). Data points belonging to each phase were separately fitted by linear least-squares regression analysis to produce the inflection point, at which point the RNAs were stoichiometrically
bound with protein.

## 2.7.2 Electron Microscopy

Electron microscopy is a useful tool to examine morphology, and it is also commonly used to characterize the size, shape, and especially the oligomerization status of proteins. Protein samples are put on copper grid and observed via an electron microscope. However, staining is necessary and otherwise the contrast would be too weak for detection. Negative staining uses heavy metal salts to enhance the contrast between the background and the object being observed. Ideally, the heavy atom salts would be uniformly placed on the copper grid except the areas with protein molecules; the protein molecules repel the heavy atom stain and thus it is call “negative” staining since the background, instead of the protein samples, is stained. Because the ability of protein molecules to deflect electron beams is weaker than that of the heavy atoms, there are more electron beams directly going through the protein. Electrons which are highly deflected are filtered out by the objective aperture, whereas the ones that go through the sample can be finally “seen” on the image. Therefore, the proteins are white and the background is dark on the final EM images.

![Diagram](http://www.snaggledworks.com)
To prepare the sample for my EM experiments, Carbon-coated copper grids (Electron Microscopy Sciences) were glow discharged at 5 mA for 1 min. 3uL of protein samples at 0.02 mg/mL was added onto the glow-discharged copper grids, washed twice with distilled water, and stained with freshly prepared 1.5% uranyl acetate solution for 1 min. Images were taken with a JEOL 2010 electron microscope (JEOL USA, Inc.) operated at 100 KeV using a magnification of 40,000X.
2.7.3 Dynamic Light Scattering

Dynamic light scattering (DLS) is a powerful tool in biochemistry and protein crystallography. It enables to measurement of the hydrodynamic radius, poly-dispersity, and the presence of protein aggregates in samples for crystallization. DLS measures the laser light scattered from dissolved macromolecules. Fluctuations in the scattering intensity can be observed because of the Brownian motion of the macromolecules. DLS measures the light scattering of a sample in short time periods and calculates the correlation between the data points. The diffusion coefficient (D) of the molecules can be calculated from the correlation function by fitting the data, and the hydrodynamic radius ($R_h$) of the macromolecule can then be calculated by:

$$R_h = \frac{kT}{6\pi \eta_0 D}$$

The molecular weight of a protein can be estimated from the measured hydrodynamic radius, using an empirical curve of known proteins and their measured hydrodynamic radii. There are empirical curves for proteins of different shapes, but unknown proteins are usually assumed to be spherical for the initial estimation.

DLS experiments were performed using a DynaPro99 Dynamic Light Scattering system from Wyatt. All of the samples were centrifuged at 22,000 x g for 30min to remove large aggregates prior to the experiments. The DLS experiments were carried out at 25°C in a buffer composed of 100mM NaCl and 25mM HEPES, pH 7.5. For each sample, 20 molecular weight readouts were calculated with each being the average of five individual measurements. The 20 readouts were further averaged to obtain the mean and standard
deviation of the protein’s molecular weight. The data analysis was done with the DYNAMICS software package.
3. Production and Characterization of the ISAV-NP

To successfully crystallize and determine the structure of a protein, it is desirable to have the target protein expressed in large quantities and to have purified the target protein to a high degree of purity. To determine the structure of the ISAV-NP, I started with the full length protein, and attempted to express it in a bacterial expression system, because of its simplicity, high yield, and relatively low cost. In this chapter, I will describe all the experiments I performed to express, purify, and characterize the full length ISAV-NP.

3.1 Cloning

The NP is encoded by an 1851bp ORF in the 3rd segment of the genomic RNA. The full length NP protein is 616aa, and the calculated molecular weight is 68 kD. I first chose to test the ISAV-NP expression in the E. coli. expression system because of its simplicity, high yield, as well as its low cost. The PCR primers (synthesized by Sigma Genosys) include two restriction enzyme digestion sites, NcoI at the 5’ end and XhoI at the 3’ end. In the forward primer, a 6X his-tag was engineered by inserting “5’-CACCACCACCACCACCAC-3’ ” after the NcoI restriction site.
The full length ISAV-NP gene (1851bp) was cloned using the pfu polymerase (Agilent Technologies, Inc.), with the 3rd segment cDNA as the amplification template. The PCR product was analyzed and confirmed by electrophoresis with a 1% agarose gel (Figure 3.1 A). The pET28b+ vector and the PCR product were double digested, ligated, and transformed into DH5α competent cells. The resulting plasmids were then obtained using a mini-prep and analyzed by double digestion to confirm that the resulting DNA fragments were of the size of the vector and ISAV-NP gene (Figure 3.1 B). Finally, the plasmid was sent for sequencing (SEQRIGHT) and the ISAV-NP gene sequence was confirmed.

3.2 Protein expression

The pET28b+ plasmid with ISAV-NP gene inserted was transformed into the *E. coli* Rosetta 2 (DE3) strain. Single colonies were inoculated into 50 mL LB Broth medium for small-scale expression tests. IPTG was added when the OD600 of the cell culture reached 0.8 to induce the ISAV-NP expression at 37°C for 6 hours (Figure 3.2 A). Most
of the expressed ISAV-NP was in the pellet fraction, and the eluted soluble protein was very limited.

Considering that salmons live in cold water, it was decided that reducing the induction temperature might be beneficial. Because the water temperature for salmons is around 13°C, and the ISAV does not replicate at a temperature above 25°C (Falk et al., 1997), I tried to express the ISAV-NP at 15°C. I did a time-series expression test of the ISAV-NP to determine the optimal duration of the induction (Figure 3.2 B). The expression experiment showed that there was no detectable expression at 1 hour. When induced for less than 5 hours, most of the expressed ISAV-NP was not soluble (in the pellet). After 7 hours, the percentage of the soluble

**Figure 3.2** Expression tests of the ISAV-NP at different temperatures. W: whole-cell lysate, S: supernatant, P: pellet, E: elution. Arrows point to the expected position of the band of the ISAV-NP in the SDS-PAGE gel. **A.** Expression at 37°C for 6 hours. Most of the expressed ISAV-NP was insoluble. **B.** Expression at 15°C. The durations of the induction are labeled on the right.
ISAV-NP out of the total protein increased, and the total expression level increased too. At 24 hours, the expression level as well as the percentage of soluble protein reached an optimal level. At the early stage of the expression, most of the expressed ISAV-NP molecules were not well-folded and thus in the insoluble form. It is possible that the ISAV-NP can be re-folded after a certain amount of time, so the low-temperature and the long-induction time are critical for the expression of soluble ISAV-NP. A 48-hour expression test has also been done and the result was comparable to that of the 24-hour induction.

### 3.3 Protein purification

4L liters of cell culture were used to express the ISAV-NP at 15°C for 24 hours. After expression, the cells were collected, lysed by sonication, and centrifuged to separate the supernatant from the pellet. The supernatant was incubated with Ni-NTA resin at 4°C for 1 hour to allow protein binding. The ISAV-NP was eluted with the Ni-NTA elution buffer containing 1M imidazole (Figure 3.3). Fractions 1-5 were collected, concentrated and dialyzed with a low-salt buffer (~200mM NaCl) for the next purification step.

![Image](image.png)

**Figure 3.3** Ni-NTA elution fractions of the ISAV-NP expressed at 15°C for 24 hours.

The ISAV-NP sample was injected into a Heparin column that was pre-equilibrated in
buffer containing 200mM NaCl. With the FPLC system, the proteins bound the Heparin column were eluted by a NaCl gradient from 0M to 1M. The ISAV-NP was eluted at a late stage (around 0.8 M NaCl), suggesting a strong binding to the Heparin column (Figure 3.4). Fractions 24-28 were collected and concentrated down to a 2mL volume for the final size exclusion chromatography step.

The Superdex-200 size-exclusion column was pre-equilibrated with a gel-filtration buffer containing 200 mM NaCl. The size-exclusion chromatography was performed with a FPLC at a flow rate of 0.7mL/min. The chromatography curve showed that there were three major peaks, each corresponding to a differently sized oligomer of the ISAV-NP based on their molecular weight shown on the SDS-PAGE. The highest peak was eluted at around 69mL, corresponding to the molecular mass of the ISAV-NP dimers when compared to the protein standards (Figure 3.5). Fractions 65-73 were collected and
concentrated for the subsequent characterization of the ISAV-NP.

**Figure 3.5** ISAV-NP elution chromatogram from a size-exclusion column. The eluted positions of three protein standards are marked by arrows. Numbers on the SDS-PAGE gel correspond to fractions at the elution volume.

### 3.4 Biochemical Characterization of the ISAV-NP

To study the crystal structure of a protein or a protein-RNA complex, it is often desired to determine whether the protein is resistant to degradation, whether it forms a single, stable molecular specie, and whether it does not aggregate upon the addition of nucleic acids. According to unpublished results from our lab, the influenza A virus NP does not behave well on some of these aspects, because it forms different oligomers that are unstable in solution. Moreover, the influenza A virus NP aggregates when DNA or RNA is added. Therefore, it is worthwhile to characterize the ISAV-NP with regards to these
aspects to see if the ISAV-NP can serve as a better model for studying NP RNA binding.

3.4.1 Degradation test

Some proteins are unstable in solution and are degraded quickly in solution by contaminating proteases, and this behavior can cause difficulties with regards to protein storage. In this experiment, purified ISAV-NP was incubated for 1 day, 4 days, and 7 days at 4°C, as well as 1 day, 4 days, 7 days, and 11 days at 22°C and 37°C respectively. SDS-PAGE analysis showed no visible degradation products (Figure 3.6).

![Degradation test of purified ISAV-NP at different temperatures.](image)

3.4.2 Oligomerization test

From Figure 3.5 we can see that the ISAV-NP exists as at least three different oligomers. Although I only collected the dimers, they could transform into other types of oligomers if fast equilibriums existed between different oligomers and monomers. Therefore, purified ISAV-NP dimers were stored at room temperature for 7 days before another size-exclusion chromatography was performed to determine if there was any change in the oligomerization status. As shown in Figure 3.7, the purified ISAV-NP remained as a homogenous dimer, which suggested that the oligomerization status of the ISAV-NP
dimer was very stable in solution at room temperature.

![Graph](image)

**Figure 3.7** Test on the oligomerization status of the ISAV-NP. The purified ISAV-NP remained as a homogenous dimer after incubating at room temperature for 7 days.

### 3.4.3 RNA-binding test

The influenza A virus NP aggregates upon the addition of RNA, because gel filtration analysis showed that the resulting NP:RNA complex of the influenza A virus eluted from the void volume of the size-exclusion chromatography. A similar experiment was performed on the purified ISAV-NP.

![Graph](image)

**Figure 3.8** Gel-filtration analysis of ISAV-NP with and without RNA addition. ISAV-NP remained as homogenous dimer upon addition of excessive RNA.
Upon the addition of excessive RNA, the ISAV-NP remained as a homogenous dimer in solution. Equal amounts of the ISAV-NP were injected for both gel-filtration analysis, and the NP:RNA peak was higher than the peak for the NP alone, presumably because of the UV absorbance of the bound RNA. There was a large RNA peak that eluted much later at around 95mL, which was likely caused by the excessive RNA.

### 3.5 Crystallization of ISAV-NP

The purified ISAV-NP was concentrated to 10mg/mL, and its crystallization was screened with about 1000 conditions using the commercial 96-well trays. Several positive conditions were identified, including:

a) 0.1M Ammonium Sulfate, 0.1M HEPES buffer (pH 7.5), and 12% (w/v) PEG 4000 (Figure 3.9 A).

b) 0.1M Lithium Sulfate, 0.1M Citrate buffer (pH 5.6), and 12% (w/v) PEG 6000

---

**Figure 3.9** Crystallization of ISAV-NP. A-C. Initial conditions identified from large-scale screen. D. Crystals obtained from manual optimization. E, F. Crystal obtained by further optimization of condition of (D) by addition extra glycerol.
c) 0.1M Ammonium Sulfate, 0.1M ADA buffer (pH 6.5), 12% (w/v) PEG 4000, and 2% (v/v) 2-propanol (Figure 3.9 C).

Each of these preliminary conditions was systematically optimized by manually setting up 24-well trays, to vary the precipitant concentration, buffer pH, salt concentration, temperature, and protein concentration. The best optimized condition was based on condition (b) described above, and the final condition was 0.95M Lithium Sulfate, 0.1M Citrate buffer (pH 5.5), and 11.5% (w/v) PEG 6000 (Figure 3.9 D). Based on this condition, an additive screen was performed with the additive screen kit (Hampton Research), and glycerol was found to be helpful in reducing the nucleation and increasing the size of the crystal (Figure 3.9 E, F). Glycerol tends to help protein to be more soluble, and thus the number of spontaneous nucleation was reduced. Therefore, if there were not as many nucleations, there would be more protein molecules packing onto the existing crystals, and thus each crystal would have a larger size. Furthermore, the concentration of glycerol in the final crystallization condition was 25%, which made the condition “cryo”, such that no more cryo-protectant was needed.

I used our Rigaku/Raxis IV++ X-ray diffraction facility to test the diffraction produced by the optimized crystals. Unfortunately, none of the crystals diffracted to a desirable resolution. The best crystal diffracted only to 8Å. Many post-crystallization treatments (Heras and Martin, 2005), such as crystal annealing, crystal dehydration, and crystal
cross-linking (described in the chapter “X-ray crystallography”), were attempted to improve the diffraction, but none of them significantly increased the resolution.

3.6 Lysine methylation of ISAV-NP

Lysine methylation was attempted to improve the diffraction of the ISAV-NP crystal. Lysine is an amino acid residue with a long, flexible size chain. The ISAV-NP has an unusually high lysine content: 7.14% (Rimstad and Mjaaland, 2002). Therefore, changes to the lysine residues may alter the surface properties of the ISAV-NP that may increase the order of protein packing, or lead to another crystal form. In fact, the reductive methylation of free amino groups on lysines has been proven to be an efficient, straightforward method to improve crystal diffraction in many cases (Kim et al., 2008; Walter et al., 2006).

Lysine methylation was performed with the JBS Methylation Kit (Jena Bioscience). After lysine methylation, the protein was again purified with size-exclusion chromatography (Figure 3.10). Because of methylation, the molecular weight of the ISAV-NP was

![Figure 3.10 Lysine methylation of ISAV-NP. As expected, the methylated ISAV-NP (red) has a large molecular weight, reflected by a smaller elution volume when compared to the native ISAV-NP (blue).](image-url)
expected to increase. Indeed, the elution volume of the methylated ISAV-NP was smaller than the native NP, as shown in Figure 3.10.

Finally, the modified ISAV-NP was concentrated to 10mg/mL and a large-scale screen was performed. However, different from the native ISAV-NP, the methylated ISAV-NP did not crystallize under any of the conditions, including those under which the native ISAV-NP could crystallize.

To summarize, although the ISAV-NP could be crystallized under a number of conditions and the shape and size of the resulting crystals were decent, the diffraction was poor. This might be caused by some internal disorder of the protein packing in the crystal, which may be led by some flexible regions of ISAV-NP. To solve the problem, sub-cloning is required to produce truncated versions of the ISAV-NP, and hopefully the flexible regions could be removed.
4. Production and Characterization of the ISAV-NP truncation mutants

Structural flexibility and instability are usually unfavorable for successful protein crystallization, because the flexible regions and polymorphism of protein may make the ordered packing of crystals difficult. Therefore, when the full length protein is resistant to crystallization or the crystals diffract poorly, there are often two ways to get around of this problem. First, one can try to crystallize the core portion of the protein and avoid the flexible parts (tails, loops). Second, one can also try to screen ligands for the protein (if not already known) and attempt to crystallize the complex, since the binding may stabilize the protein or change its conformation in a way that may facilitate crystallization.

4.1 Construct design

To obtain a more compact NP truncation mutant that would produce better NP crystals, I employed different techniques, including limited proteolysis, sequence analysis, and secondary structure analysis, to identify suitable NP constructs.

4.1.1 Limited proteolysis

To determine the flexible regions on a protein, limited proteolysis is often used because those flexible regions are usually more exposed and more accessible to proteases. From this experiment, I was looking for a dominant “sub-protein” to use it as an alternative
target for protein crystallization. The flexible regions may not be visible on the electron
density map anyways, and their existence may cause crystal packing disorder.

Trypsin and chymotrypsin were used for this experiment. All digestion experiments were
performed at 37°C in 50mM Tris pH7.5, 200mM NaCl, 10% glycerol (v/v), and 5 mM
β-mercaptoethanol. First, the purified ISAV-NP was digested by proteases at various
concentrations, ranging from 1:50 to 1:20000 mass ratios of protease:ISAV-NP, for 30
minutes (Figure 4.1). It was determined the best ratio was 1:500 (mass ratio) between

![Limited proteolysis of the ISAV-NP by trypsin and chymotrypsin at different protease:protein ratios. Ratios are labeled on top. The position of the full length protein is marked by arrow.](image1)

**Figure 4.1** Limited proteolysis of the ISAV-NP by trypsin and chymotrypsin at different protease:protein ratios. Ratios are labeled on top. The position of the full length protein is marked by arrow.

![Time-series digestion profile of the ISAV-NP by trypsin at 37°C for 30 minutes with a protease:protein mass ratio of 1:500. Arrow points to the best duration of the digestion.](image2)

**Figure 4.2** Time-series digestion profile of the ISAV-NP by trypsin at 37°C for 30 minutes with a protease:protein mass ratio of 1:500. Arrow points to the best duration of the digestion.
trypsin and the ISAV-NP for a 30 minute digestion. Second, the best time length of the digestion at the 1:500 ratio was determined (Figure 4.2), and it turned out that the digestion was best at 40 minutes because there was a clear band of “sub-protein” with almost no contaminating protein.

After determining the best conditions for the trypsin digestion, I purified the full length ISAV-NP and performed digestion at such condition followed by purification with the Heparin column and the Superdex 200 size-exclusion column with the FPLC. Because the full length ISAV-NP has an N-terminal his-tag, and a western-blot with an anti-his antibody showed that the digested ISAV-NP did not have his-tag, I determined the N-terminus was cleaved during the digestion. The purified ISAV-NP truncation was sent for N-terminal sequencing (Tufts core facilities, Boston MA) which suggested the N-terminal 188 residue was cleaved. Based on the estimation of the molecular weight by mass spectrometer (Tufts core facilities, Boston MA), I found that the C terminus was intact. Therefore, this truncation was name ΔN189.

Beside this limited-proteolysis experiment, I accidentally identified another ISAV-NP truncation during the purification of the full length ISAV-NP, possibly due to not adding sufficient protease inhibitor. This accidental proteolytic truncation was name ΔN111.

Figure 4.3 An ISAV-NP truncation ΔN111 obtained by accidental proteolytic cleavage.
ISAV-NP truncation was cleaved at a different position from that of the trypsin-digested protein, because the molecular weights were different (Figure 4.3). Like the trypsin-digested protein, this truncated ISAV-NP was purified, and analyzed by N-terminal sequencing and mass spec. It was also an N-terminal truncation of the ISAV-NP, without the N-terminal 111 residues. It was named ΔN111.

### 4.1.2 Primary sequence analysis

Analysis of a protein's amino acid sequence can also help in locating flexible regions. Because the NP structure of the Influenza A virus showed that the first 20 amino acids at the N terminus are highly flexible (Ye et al., 2006), I decided to delete a few amino acids on the N' termini of the ISAV-NP and see if this helps in the crystallization of the ISAV-NP, assuming that the ISAV-NP resembles the influenza A virus NP. Finally I decided to deleted the first 26 residues and this truncation was named ΔN26.

On the other hand, I found that 12 out of the 16 C-terminal amino acids of the ISAV-NP are acidic ones (-EIEFDEDDEEEEDIDI-) which are likely to be flexible in solution. Therefore I decided to delete this C-terminal tail region and see whether the quality of the resulting crystals could be improved. This truncation was name ΔC16.

### 4.1.3 Secondary structure analysis

According to the crystal structure of the influenza A virus NP, the first 23 amino acids at the N terminus form a random coil. It is possible that the N terminus of the ISAV-NP is
also flexible and may have some adverse effects on crystal packing. Because the primary sequence alignment of the ISAV and Influenza A virus NPs showed little similarity (this does not necessarily mean that their structures will be totally different), I decided to do a secondary structure prediction to search for possible flexible regions at the N terminus of the ISAV NP. As shown, the online software PREDICTPROTEIN showed four possible α-helices each starting at amino acids #39, #50, #64, and #84 (Figure 4.4). Therefore, assuming that the α-helices are relatively rigid structures, I have designed four truncations of the ISAV-NP starting from the beginning of each of these four α-helices, so as to reduce the flexibility at the N-terminal end of the protein. These four N-terminal truncations were named ΔN39, ΔN50, ΔN64, and ΔN84, respectively.

Figure 4.4 Secondary structure prediction of a section of the ISAV-NP. Amino acids are numbered on top. The letters below the amino acids show the secondary structure prediction, with H=helix, E=extended (sheet), Blank=other (loop). The arrows on the top show the truncated positions in the four N-terminal truncation mutants, ΔN39, ΔN50, ΔN64, and ΔN84.
4.2 Expression, purification and crystallization of the ISAV-NP truncations

Each ISAV-NP truncation was cloned in the same way as the full length ISAV-NP. The expression level and protein solubility were usually tested at both 37°C and 15°C. If the protein was expressed and the solubility was good, purification was performed in a fashion similar to that of the full-length ISAV-NP, since it is likely that the truncations share similar biochemical and biophysical properties with the full length ISAV-NP. Eight plasmids were constructed to express the aforementioned eight truncations of the ISAV-NP, with an N-terminal his-tag.

a. ΔN189

The ISAV-NPΔN189 was designed according to the trypsin digestion experiment discussed above. Expression of the ISAV-NPΔN189 was tested with 1mM IPTG induction at 37°C for 6 hours and at 15°C for 24 hours. When induced at 37°C, most of the expressed ISAV-NPΔN189 was insoluble and in the pellet fraction (Figure 4.5). When the induction temperature was lowered to 15°C, there was no detectable expression (Figure 4.5). I also tried lowering the IPTG concentration, but the amount of soluble ISAV-NPΔN189 did not increase.

Figure 4.5 Expression of the ISAV-NP ΔN189 at 37°C and 15°C. W: whole-cell lysate, S: supernatant, P: pellet. The arrow points to the expected position of truncation, based on its molecular
Because of the low yield of soluble protein, I did not continue to purify and try crystallization screens on the ISAV-NPΔN189.

b. ΔN111

The ISAV-NP ΔN111 was expressed well in *E. coli* by 1mM IPTG induction at 15 Celsius for 24 hours (Figure 4.6 A), and the solubility was also comparable to the full length ISAV-NP. Soluble ISAV-NP ΔN111 was collected from the supernatant and purified with Ni-NTA agarose resin, followed by the Heparin column, ion-exchange column (HP SP column), and finally by the size-exclusion column in the refrigerated FPLC system (Figure 4.6 B-D).

To characterize the oligomerization status of ISAV-NP ΔN111, I used negative-staining electron microscopy to observe the protein molecules. The ISAV-NP ΔN111 was stained at the concentration of 0.02mg/ml with uranyl formate and observed at 40,000 times magnification. EM images showed that most ISAV-NP ΔN111 are dimerized, consistent with its elution profile from the size-exclusion column (Figure 4.6 D, I).

Prior to the crystallization test, the ISAV-NP ΔN111 was concentration to 10mg/mL. Crystallization screens were performed with the “Hydra II Plus One” robot followed by manual optimization. Crystals from several different conditions were obtained (Figure 4.6 E-H) and tested with our Rigaku/Raxis IV++ X-ray diffraction facility. Unfortunately it showed little improvement compared to the diffraction of full length ISAV-NP crystals.
Figure 4.6 Expression, purification, crystallization and characterization of ISAV-NP ΔN111. 

A. SDS-PAGE to analyze the Ni-NTA purification. W: whole-cell lysate, S: supernatant, P: pellet, FT: flow through. The numbers correspond to the number of eluted fractions. B-D. Purification with the Heparin column (B), HP SP column (C) and Superdex-200 size-exclusion column (D). The SDS-PAGE shows the purity of the protein, and the numbers on top correspond to the fraction numbers. In (D), the ISAV-NP ΔN111 peak corresponds to the dimers is pointed with two arrows. E-H. Crystal images of ISAV-NP ΔN111. I. Negative-staining EM to characterize the oligomerization status of ISAV-NP ΔN111. Dimers were highlighted by white circles.
c. ΔN26

The expression of the ISAV-NPΔN26 was tested with 1mM IPTG induction at 37°C for 6 hours and at 15°C for 24 hours. At 37°C, the ISAV-NPΔN26 did not express. At 15°C, the ISAV-NPΔN26 expression might be detected, but most of the protein was insoluble (Figure 4.7). Because of the low yield of soluble protein, I did not continue to purify and try crystallization screens on the ISAV-NPΔN26.

d. ΔC16

ISAV-NPΔC16 was expressed in *E. coli* with 1mM IPTG induction at 15°C for 24 hours. The expression level and protein solubility of ISAV-NPΔC16 was even better than the full length ISAV-NP. Soluble ISAV-NPΔC16 was collected in the supernatant and purified with Ni-NTA agarose, followed by a Heparin column, an ion-exchange column (HP SP) and a finally size-exclusion column (Superdex-200) in a refrigerated FPLC system (Figure 4.8). The protein was concentrated to 10mg/mL and a large-scale crystallization screen was performed. Unluckily this ISAV-NPΔC16 did not crystallize under those ~1000 conditions from the screen.
Figure 4.8 Expression, purification, crystallization and characterization of the ISAV-NP ΔC16.  
A. SDS-PAGE gel to analyze the Ni-NTA purification. W: whole-cell lysate, S: supernatant, P: pellet, Ft: flow through. The numbers correspond to the number of eluted fractions.  
B-D. Purifications with a Heparin column (B), HP SP column (C) and a Superdex-200 size-exclusion column (D). The SDS-PAGE gel shows that the purity of the protein after each step, and the numbers on top corresponds to the fraction numbers.
e. ΔN39, ΔN50, ΔN64, and ΔN84

All of these four truncated protein were designed according to the secondary structure prediction of the ISAV-NP, and they were all successfully expressed in *E. coli.* under similar conditions as the full length ISAV-NP. Interestingly, the yields of soluble proteins per liter were slightly different and following this trend: ΔN39 < ΔN50 < ΔN64 < ΔN84 < full length ISAV. All four truncations were purified with Ni-NTA resin, a Heparin column, an ion-exchange column and finally a size-exclusion column (Figure 4.9). After size-exclusion chromatography, SDS-PAGE gel analysis was performed to confirm the purity. From the SDS-PAGE gel analysis, I found that ΔN39 and ΔN84 were very homogeneous (with a single band), whereas ΔN50 and ΔN64 contained several proteolytic products suggesting that ΔN50 and ΔN64 have some small flexible regions.

![Figure 4.9 Size-exclusion chromatography of ΔN39 (A), ΔN50 (B), ΔN64 (C), and ΔN84 (D). The arrows point to the peaks corresponding to the protein collected. The SDS-PAGE gel shows the purity of each of the proteins after purification.](image-url)
that are susceptible to protease digestion.

All four proteins were subjected to an initial screen for crystallization conditions. ΔN50, ΔN64, and ΔN84 were crystallized, and each of the crystallization conditions were manually optimized by varying the concentration of salt, the concentration of precipitant, pH, the type of buffer, the type of salt, temperature, etc. Finally ΔN84 formed larger and better-looking crystals than ΔN50 and ΔN64 (Figure 4.10). Crystals of the three truncations were tested in our in-house x-ray generator, but none of them gave satisfactory diffraction. The best crystal diffracted to 7-8 Å only.
4.3 Summary of the ISAV-NP truncations

For most crystallographers, truncation is usually one of the first things to try if a protein fails to crystallize or if the diffraction is poor. The reasoning behind this is that rational truncations can remove “flexible” portions of the protein and make it more “crystallizable”, or improve the packing of the crystals if crystallization is already achieved. Usually truncations are made by deleting the C terminus and/or the N terminus, since the middle part is usually the core domain with important biological functions.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Expression</th>
<th>Solubility</th>
<th>Yield</th>
<th>Crystallizability</th>
<th>Diffraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL</td>
<td>Yes(37°C/15°C)</td>
<td>Good</td>
<td>2mg/L</td>
<td>Yes</td>
<td>7-8Å</td>
</tr>
<tr>
<td>ΔN189</td>
<td>Yes(37°C)</td>
<td>Poor</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>ΔN111</td>
<td>Yes(37°C/15°C)</td>
<td>Good</td>
<td>2mg/L</td>
<td>Yes</td>
<td>10Å</td>
</tr>
<tr>
<td>ΔN26</td>
<td>Yes(15°C)</td>
<td>Poor</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>ΔC16</td>
<td>Yes(37°C/15°C)</td>
<td>Excellent</td>
<td>5mg/L</td>
<td>No</td>
<td>N/A</td>
</tr>
<tr>
<td>ΔN39</td>
<td>Yes(15°C)</td>
<td>Good</td>
<td>0.5mg/L</td>
<td>No</td>
<td>N/A</td>
</tr>
<tr>
<td>ΔN50</td>
<td>Yes(15°C)</td>
<td>Good</td>
<td>0.5mg/L</td>
<td>Yes</td>
<td>7-8Å</td>
</tr>
<tr>
<td>ΔN64</td>
<td>Yes(15°C)</td>
<td>Good</td>
<td>0.5mg/L</td>
<td>Yes</td>
<td>7-8Å</td>
</tr>
<tr>
<td>ΔN84</td>
<td>Yes(15°C)</td>
<td>Good</td>
<td>0.5mg/L</td>
<td>Yes</td>
<td>7-8Å</td>
</tr>
</tbody>
</table>

Table 4.1 Summary of ISAV-NP full length protein and all ISAV-NP truncations.

I made eight truncations to the ISAV-NP (Table 4.1), seven at the N termini and one at the C terminus. Six out of the eight proved soluble and five of them could be crystallized. Unfortunately, none of them really improved the quality of the crystal diffraction. I think the most plausible explanation for these findings is that the core domain, at least a part of it, is very flexible, although this is not often seen in proteins. If so, truncations at both termini of the ISAV-NP would do little to improve crystal quality and other approaches were employed to make the protein more rigid.
5. The assembly of the ISAV-NP and nucleic acid complexes

Structural flexibility and instability are usually unfavorable for successful crystallization. Since the ISAV-NP is a DNA/RNA binding protein, adding ligands (DNA/RNA) prior to crystallization could help to stabilize the ISAV-NP. When I started the ISAV-NP project, I did not directly try to crystallize the complex of ISAV-NP:DNA, because the homogeneity of the sample, which was one of the most important factors for crystallization, could be much harder to control in co-crystallization than crystallizing the ISAV-NP alone. However, since neither the full length ISAV-NP nor its truncations gave well-diffracting crystals, it was then necessary to study the binding of the ISAV-NP to DNA, and attempt the co-crystallization of the protein-nucleic acid complex.

5.1 Strategy for forming the complex

There were several candidate proteins to choose from: the full length ISAV-NP, ΔN111, ΔC16, and all the other soluble ISAV-NP truncations. Also, I needed to decide on the length of RNA to use to form the complex with the protein. In addition, I needed to decide on the ratio of protein and RNA, the buffer condition (salt, pH, and etc.), and the temperature. There would be too many conditions to test if I were to try every possible combination of the above factors for producing the NP:RNA complex. Therefore, it was beneficial to first consider the conditions that favor the formation of the NP:RNA complex as well as the subsequent crystallization.
5.1.1 Choice of protein

There were a number of ISAV-NP proteins that could serve as the candidate for forming the complex. From the previous information, soluble proteins included the full length ISAV-NP, ISAV-NP ΔN111, ΔC16, ΔN39, ΔN50, ΔN64, and ΔN84 (Table 5.1).

<table>
<thead>
<tr>
<th>Protein</th>
<th>Expression</th>
<th>Solubility</th>
<th>Yield</th>
<th>Crystallizability</th>
<th>Diffraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL</td>
<td>Yes(37°C/15°C)</td>
<td>Good</td>
<td>2mg/L</td>
<td>Yes</td>
<td>7-8Å</td>
</tr>
<tr>
<td>ΔN189</td>
<td>Yes(37°C)</td>
<td>Poor</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>ΔN111</td>
<td>Yes(37°C/15°C)</td>
<td>Good</td>
<td>2mg/L</td>
<td>Yes</td>
<td>10Å</td>
</tr>
<tr>
<td>ΔN26</td>
<td>Yes(15°C)</td>
<td>Poor</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>ΔC16</td>
<td>Yes(37°C/15°C)</td>
<td>Excellent</td>
<td>5mg/L</td>
<td>No</td>
<td>N/A</td>
</tr>
<tr>
<td>ΔN39</td>
<td>Yes(15°C)</td>
<td>Good</td>
<td>0.5mg/L</td>
<td>No</td>
<td>N/A</td>
</tr>
<tr>
<td>ΔN50</td>
<td>Yes(15°C)</td>
<td>Good</td>
<td>0.5mg/L</td>
<td>Yes</td>
<td>7-8Å</td>
</tr>
<tr>
<td>ΔN64</td>
<td>Yes(15°C)</td>
<td>Good</td>
<td>0.5mg/L</td>
<td>Yes</td>
<td>7-8Å</td>
</tr>
<tr>
<td>ΔN84</td>
<td>Yes(15°C)</td>
<td>Good</td>
<td>0.5mg/L</td>
<td>Yes</td>
<td>7-8Å</td>
</tr>
</tbody>
</table>

*Table 5.1 Summary of ISAV-NP full length protein and all ISAV-NP truncations.*

First, for the efficient screening of crystallization conditions, it was desirable to have a relatively high yield of the protein such that more conditions could be screened at a time. According to the above table, the full length protein, ISAV-NP ΔN111, and ΔC16 had a relatively high yield. Therefore, I focused on choosing from these three proteins.

Second, to retrieve more structural information from the final structure, it was desirable to crystallize the full length protein or a truncation that retained as much of the protein sequence as possible. Among the three candidates, ΔN111 was significantly shorter than the others, so ΔN111 was not my first choice for producing the NP:RNA complex.
Third, to crystallize the NP:RNA complex, it was important to have the complex stable in solution, which required the protein to have a strong affinity to RNA.

<table>
<thead>
<tr>
<th></th>
<th>20nt-RNA</th>
<th>24nt-RNA</th>
<th>28nt-RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISA V NP</td>
<td>234.6nM</td>
<td>24.5nM</td>
<td>9.2nM</td>
</tr>
<tr>
<td>ΔC16</td>
<td>25.5nM</td>
<td>4.3nM</td>
<td>2.2nM</td>
</tr>
</tbody>
</table>

*Table 5.2* Fluorescence Polarization tests to compare the dissociation constants (K_d) of full length ISA V NP and ISA V NP ΔC16 when binding to different lengths of RNA.

Therefore, it was necessary to compare the RNA binding affinity of the full length ISA V NP and the ISA V NP ΔC16. Fluorescence polarization was employed to compare the RNA-binding affinities of the two proteins. As shown, the dissociation constants (K_d) of ΔC16 is much smaller (approximately 10-fold) than that of full length ISA V NP, showing a much stronger binding of ΔC16 with RNA (Table 5.2). Therefore I decided to start with ΔC16 instead of the full length protein for the co-crystallization experiments.

In addition, as I mentioned before, the C terminus tail of ISA V NP is very acidic; 12 out of 16 amino acids are either glutamic acids or aspartic acids. I hypothesize that this highly-negatively charged C-terminal tail may inhibit the protein-RNA binding by interacting with the RNA-binding groove which is highly positively charged. In other words, the C terminal tail may impede the protein-RNA binding in a competitive manner. This acidic C-terminal tail is actually conserved in the Influenza A NP and the Influenza B NP, so it is likely to play an important biological role for all orthomyxoviruses.
5.1.2 Choice of RNA

On the basis of genome-wide stoichiometry calculations, it was estimated that there are ~24 nts of RNA for each NP molecule in influenza A virus (Compans et al., 1972; Martin-Benito et al., 2001; Ortega et al., 2000). However, that was based on the calculated ratio of the length of genomic RNA to the number of NPs bound. If there were NP-free RNA regions, there would be less than 24nt of RNA for each NP molecule. Furthermore, the footprint of RNA for each NP would be even less on considering the gaps in between the NPs. Therefore, I decided to start with testing RNAs shorter than 24nt.

Moreover, the longer the RNA is, the greater the chance that two protein molecules might bind to the same RNA molecule. Undoubtedly, these 2:1 complexes would reduce the homogeneity of the sample and adversely affect crystallization. Therefore, the optimal length for co-crystallization was probably not the one that ensured the strongest binding, but the minimal length that still sufficed the protein-RNA binding.

Therefore, I tried to make complexes by mixing ΔC16 with different lengths of RNAs, and analyzing the complex by size-exclusion chromatography. I purified the ISAV ΔC16 with Ni-NTA resin, the Heparin column, an ion-exchange column, and finally a size-exclusion column as previously described. I added DNA oligos of different lengths to the protein sample and ran the size-exclusion column again to purify the complexes. 15-polydT, 18-polydT, and 20-polydT were tested with the ISAV ΔC16 in 200mM NaCl
(Figure 5.1). Because binding to nucleic acids increases the molecular weight, successful

**Figure 5.1** Chromatographic curves of different proteins (or protein:DNA complexes) from the GE Superdex 200 column. From A to D: ISAV ΔC16 (brown), ISAV ΔC16 + polydC15 (green), ISAV ΔC16 + polydC18 (blue), ISAV ΔC16 + polydC20 (red). E: Superimposition of all four curves at 150ml-190ml region to show the relative positions of the peaks. From A-D: red arrows on the top point to the peaks corresponding to the protein complex collected, and blue arrows on the top point to the peak of excessive DNAs.
complex formation would shift the peak to the left-hand side compared to the peak of ISAVΔC16 only. After superimposing the four peaks, I found that the addition of polydC20 (red) increased the molecular weight the most. The addition of polydC18 (blue) also increased the molecular weight, since both the red and blue curves have been shifted to the left compared to the brown curve (protein only). However, the addition of polydC15 failed to shift the curve, suggesting the polydC15 may be too short for the protein to bind. Therefore, I choose dC18 for the following co-crystallization experiment.

5.1.3 Other parameters

First, to ensure the homogeneity of the sample, I decided to purify the complex with size-exclusion chromatography before crystallization. Therefore, the ratio of the nucleic acids to the protein were not critical since the excessive nucleic would be removed by the purification. To favor the binding, the amount of nucleic acids was usually in excess. In practice, the nucleic acids and protein were mixed at a molar ratio of 1.5:1 for the complex formation. Another consideration to let the nucleic acids be in excess instead of the protein was that it would be easier to separate the complex from nucleic acids than from apo protein with the size-exclusion chromatography, due to the molecular weight differences.

To favor binding, the mixture of nucleic acids and protein were incubated at room
temperature for 1 hour before being applied to the size-exclusion column. The salt concentration of the gel-filtration buffer used in size-exclusion chromatography was kept as low as possible to favor the stability of the complex of protein and nucleic acid. For the ISAV-NP ΔC16, 150mM was chosen as the final NaCl concentration based on a solubility test.

The freshness of the sample was also important for crystallization. Once the complex was purified, centrifugation was immediately performed to concentrate the sample to 10mg/mL. One thing to note here was that determination of the complex concentration by UV spectrometer was not possible. Instead, the Bradford assay was employed to estimate the concentration. Crystallization screening was performed when the sample was as fresh as possible.

5.2 Co-crystallization of ISAV-NPΔC16 and nucleic acids

Crystals of the ISAV-NP ΔC16 were obtained from a protein-ssDNA complex containing the oligo dC18. The presence of DNA was required for crystal formation, since ISAV-NP ΔC16 did not crystallize under similar conditions. The best crystals were grown at 10°C for two weeks using the sitting-drop vapor diffusion method with the drop made of equal volumes of protein solution (10mg/ml in 200mM NaCl, 10% glycerol, and 50 mM Tris, pH7.5) and well solution (containing 0.8M lithium chloride, 0.1M sodium citrate, pH 4.0, and 20% PEG 6000). To improve the crystal diffraction, an additional dehydration step was performed by transferring crystals to a mother liquor
containing 0.8M lithium chloride, 0.1M sodium citrate, pH 4.0, 9% PEG 6000 and 20% PEG 400. After 24 hours, the crystals were directly frozen in liquid nitrogen. The crystals diffracted to $\sim$4Å with our home X-ray source, which was promising.

To solve the phase problem (described in the chapter “X-ray crystallography”), I chose to use single-wavelength anomalous dispersion (MAD), which required the preparation of SeMet protein by replacing the methionine with selenomethionine (SeMet). There are 27 methionines in the ISAV-NP ΔC16. I expressed the SeMet-protein in M9 media, and purified the protein with the same procedure as the wild-type protein. The protein was successfully crystallized under similar conditions as the native protein. To optimize the crystallization, seeding was used to obtain fewer and larger crystals, followed by crystal dehydration to obtain tighter crystal packing. The diamond-like crystals grew to 0.4mm in the longest dimension (Figure 5.2A), and diffracted to around 3.1Å in our home X-ray source (Figure 5.2B).

![Figure 5.2A](image)

**Figure 5.2A.** Crystals of the complex of SeMet-ISAV ΔC16:polydC18 after optimization.

**Figure 5.2B.** $\sim$3.1Å diffraction image of the crystal tested at our in-house X-ray diffraction facility.
6. Data collection and structure determination

Although crystallization is usually the rate-limiting step, to successfully determine a structure, one also needs to be familiar with other techniques, such as post-crystallization manipulation, data collection and processing, as well as model building and refinement. In this chapter, I will describe the process of data collection and structure determination of the ISAV-NP. Problems I encountered will be presented and the solutions will be discussed.

6.1 Improve ISAV-NP crystal by crystal manipulation

The crystallization conditions from the initial screen usually need to be optimized to get satisfactory X-ray diffraction. In this thesis project, after extensive manual optimization, the diffraction was improved to 3.0Å at the home source. However, the crystals were very fragile and usually cracked when they were transferred to the cryo-protectant. Seeding was used in combination with the crystal dehydration method to increase the robustness of the crystals, as well as the resolution. Finally, the resolution was extended to 2.7Å. The crystals were initially grown at 10°C for two weeks using the sitting-drop vapor diffusion method with the drop made of equal volume of protein solution (10mg/ml in 200mM NaCl, 10% glycerol, and 50 mM Tris, pH7.5) and well solution (containing 0.8M lithium chloride, 0.1M sodium citrate, pH 4.0, and 20% PEG 6000). Seeding was performed with seeding beads (Hampton research) to obtain fewer and
bigger crystals. To further improve the crystal quality, an additional dehydration step was
performed by transferring crystals to a mother liquor containing 0.8M lithium chloride,
0.1M sodium citrate, pH 4.0, 9% PEG 6000 and 20% PEG 400. After 24 hours, the
crystals were directly frozen in liquid nitrogen. The crystals finally diffracted to 2.7Å at
the beamline 4.2.2 at the Advanced Light Source, Berkeley.

6.2 Data collection

X-ray diffraction tests were usually first performed with our in-house Rigaku/Raxis
IV++ X-ray diffraction facility. For higher intensity and monochromaticity, we used
synchrotron X-ray sources. The home source has its advantage because it is more
convenient and readily available, but with synchrotron X-ray sources one can usually
collect one full data set within 1 hour since a much shorter exposure is required for each
image, and the resolution is usually improved. The data for this thesis project was
collected at the Beamline 4.2.2 at the Advanced Light Source, Berkeley, with a
wave-length of 0.97895Å at 100K, and the diffraction images were detected and
recorded with the "NOIR-1" MBC system. Crystals were prescreened at the home source
and stored in liquid nitrogen before shipping to the synchrotron X-ray source for data
collection. The data collection was performed with an oscillation angle of 0.5°, to avoid
overlapping. Because the radiation damage to my crystals was significant, collection of a
full data set of 180 degrees was not possible. After indexing the initial snapshot image,
the crystal’s space group was determined to be P422 which has relatively high symmetry,
and therefore a large oscillation range was not necessary. In the end, my crystal was
collected with a 170mm distance, an oscillation angle of 0.5 degree, an oscillation range of 75 degree, and an exposure time of 4 seconds per frame.

6.3 Structure determination

To determine a protein structure, one needs the aid of several software packages such as HKL2000, Phenix, and COOT. For this project, the HKL2000 suite was used for the analysis of X-ray diffraction data collected from single crystals, including visualization of the diffraction images/patterns, parameter refinement, data reduction and integration, merging and scaling the intensity information, and generating the final statistics (Otwinowski and Minor, 1997). Next, Autosol from the Phenix program was used for the determination of heavy-atom sites, the calculation of experimental phases, the density modification, and the model-building. Finally, the structure model was manually built and modified with COOT, and validated with Phenix.validate.

a. HKL2000

I started indexing with HKL2000. For my ISAV-NP crystal, the Primitive tetragonal Bravais lattice was chosen, with the unit cell dimensions of \( a = b = 113.06 \, \text{Å}, \ c = 103.50 \, \text{Å} \), and \( \alpha = \beta = \gamma = 90^\circ \) (Figure 6.1). The predicted reflections agreed with the actual ones on the images (Figure 6.2), and the \( \chi^2 \) values (Figure 6.3) were good (green and less than 1), indicating good data quality. Therefore, I was able to proceed to the next steps, the integration and scaling.
Figure 6.1 Bravais lattice table generated with HKL2000. Possible Bravais lattice (distortion < 1%) are highlighted in green. The Bravais lattices are list with the ones of higher symmetries on top and the ones with lower symmetries at bottom.
Figure 6.2 Reflection image after refinement shows the most of predicted spots agree with the actual ones.

<table>
<thead>
<tr>
<th>Refinement Information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space Group: P4</strong></td>
</tr>
<tr>
<td>Resolution: 50.00 - 2.21</td>
</tr>
<tr>
<td>Positional: 1810</td>
</tr>
<tr>
<td>Partiality: 3138</td>
</tr>
<tr>
<td>X Beam: 79.883</td>
</tr>
<tr>
<td>a: 112.95</td>
</tr>
<tr>
<td>(\alpha): 90.00</td>
</tr>
<tr>
<td>Crystal Rotation X: -10.710</td>
</tr>
<tr>
<td>Crystal Rotation Y: 15.281</td>
</tr>
<tr>
<td>Crystal Rotation Z: 62.277</td>
</tr>
<tr>
<td>Detector Rotation X: 0.124</td>
</tr>
<tr>
<td>Detector Rotation Y: 0.485</td>
</tr>
<tr>
<td>Detector Rotation Z:</td>
</tr>
<tr>
<td>Crossline X: 0.027</td>
</tr>
<tr>
<td>Crossline Y: 0.027</td>
</tr>
<tr>
<td>Crossline XY: -0.096</td>
</tr>
<tr>
<td>Distance: 169.670</td>
</tr>
<tr>
<td>Mosakity: 0.736</td>
</tr>
</tbody>
</table>

Figure 6.3 The final refinement information from HKL2000. The \(\chi^2\) values are shown in the top panel. The second column in the bottom panel is the error terms which should converge to zero after refinement.
After scaling, a .sca file was generated which contained all of the structural data from the X-ray diffraction. The .log file was examined and the statistics were listed in Table 6.1, 6.2, and 6.3. Overall the data quality was very good. The average Rmerge was 12.0 %, and the Rmerge value for the highest resolution shell (2.75 - 2.7Å) was 0.458, which showed very decent quality (Table 6.1). For all resolution bins, the I/σ values were over 2.0 (Table 6.1), the completeness values were over 99% (Table 6.2), and redundancy values were over 4.9 (Table 6.3), all of which showed the data quality was good.

<table>
<thead>
<tr>
<th>Shell Lower Limit (Å)</th>
<th>Upper Limit (Å)</th>
<th>Average I</th>
<th>Mean I/error (σ)</th>
<th>Mean stat. Chi**2</th>
<th>R-fac</th>
<th>R-fac Linear Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>50.00</td>
<td>7.32</td>
<td>11409.7</td>
<td>995.5</td>
<td>118.9</td>
<td>0.939</td>
<td>0.071</td>
</tr>
<tr>
<td>7.32</td>
<td>5.81</td>
<td>4358.1</td>
<td>398.3</td>
<td>104.0</td>
<td>0.990</td>
<td>0.094</td>
</tr>
<tr>
<td>5.81</td>
<td>5.08</td>
<td>4051.8</td>
<td>366.0</td>
<td>110.9</td>
<td>0.994</td>
<td>0.102</td>
</tr>
<tr>
<td>5.08</td>
<td>4.62</td>
<td>4976.2</td>
<td>476.3</td>
<td>125.7</td>
<td>1.009</td>
<td>0.104</td>
</tr>
<tr>
<td>4.62</td>
<td>4.29</td>
<td>4797.7</td>
<td>433.6</td>
<td>117.7</td>
<td>1.017</td>
<td>0.106</td>
</tr>
<tr>
<td>4.29</td>
<td>4.03</td>
<td>4272.9</td>
<td>401.1</td>
<td>107.4</td>
<td>1.034</td>
<td>0.108</td>
</tr>
<tr>
<td>4.03</td>
<td>3.83</td>
<td>3524.2</td>
<td>345.5</td>
<td>121.1</td>
<td>1.176</td>
<td>0.123</td>
</tr>
<tr>
<td>3.83</td>
<td>3.66</td>
<td>2802.8</td>
<td>281.6</td>
<td>166.2</td>
<td>1.133</td>
<td>0.131</td>
</tr>
<tr>
<td>3.66</td>
<td>3.52</td>
<td>2301.9</td>
<td>234.0</td>
<td>160.5</td>
<td>1.077</td>
<td>0.144</td>
</tr>
<tr>
<td>3.52</td>
<td>3.40</td>
<td>1876.8</td>
<td>201.2</td>
<td>93.6</td>
<td>1.033</td>
<td>0.153</td>
</tr>
<tr>
<td>3.40</td>
<td>3.30</td>
<td>1537.0</td>
<td>171.5</td>
<td>87.9</td>
<td>0.927</td>
<td>0.160</td>
</tr>
<tr>
<td>3.30</td>
<td>3.20</td>
<td>1074.5</td>
<td>135.1</td>
<td>82.0</td>
<td>0.987</td>
<td>0.187</td>
</tr>
<tr>
<td>3.20</td>
<td>3.12</td>
<td>898.7</td>
<td>124.2</td>
<td>81.2</td>
<td>0.867</td>
<td>0.216</td>
</tr>
<tr>
<td>3.12</td>
<td>3.04</td>
<td>698.2</td>
<td>169.8</td>
<td>79.2</td>
<td>0.858</td>
<td>0.249</td>
</tr>
<tr>
<td>3.04</td>
<td>2.97</td>
<td>585.0</td>
<td>102.6</td>
<td>77.5</td>
<td>0.760</td>
<td>0.270</td>
</tr>
<tr>
<td>2.97</td>
<td>2.91</td>
<td>444.6</td>
<td>94.6</td>
<td>77.5</td>
<td>0.765</td>
<td>0.333</td>
</tr>
<tr>
<td>2.91</td>
<td>2.85</td>
<td>380.6</td>
<td>91.2</td>
<td>76.6</td>
<td>0.773</td>
<td>0.376</td>
</tr>
<tr>
<td>2.85</td>
<td>2.80</td>
<td>329.7</td>
<td>88.1</td>
<td>76.6</td>
<td>0.739</td>
<td>0.410</td>
</tr>
<tr>
<td>2.80</td>
<td>2.75</td>
<td>309.8</td>
<td>90.2</td>
<td>79.4</td>
<td>0.760</td>
<td>0.434</td>
</tr>
<tr>
<td>2.75</td>
<td>2.70</td>
<td>272.2</td>
<td>92.5</td>
<td>83.0</td>
<td>0.734</td>
<td>0.458</td>
</tr>
</tbody>
</table>

Table 6.1 The R_{merge} and I/σ values for each resolution bin. The R\textsubscript{merge} is listed in the second column from the right, and the intensities of I (signal) and σ(noise) are listed in the third and fourth column from left.
Table 6.3 The completeness of each resolution bin. The total completeness is shown in the first column from the right.

<table>
<thead>
<tr>
<th>Shell</th>
<th>I/Sigma in resolution shells:</th>
<th>Lower</th>
<th>Upper</th>
<th>% of</th>
<th>of</th>
<th>reflections with I / Sigma less than 20</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>limit</td>
<td>limit</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>50.00</td>
<td>7.32</td>
<td>1.1</td>
<td>1.7</td>
<td>2.2</td>
<td>2.5</td>
<td>4.1</td>
<td>33.1</td>
</tr>
<tr>
<td>7.32</td>
<td>5.81</td>
<td>1.6</td>
<td>2.9</td>
<td>4.5</td>
<td>5.7</td>
<td>8.5</td>
<td>37.4</td>
</tr>
<tr>
<td>5.81</td>
<td>5.08</td>
<td>3.4</td>
<td>5.4</td>
<td>7.0</td>
<td>7.8</td>
<td>11.0</td>
<td>41.0</td>
</tr>
<tr>
<td>5.08</td>
<td>4.62</td>
<td>3.6</td>
<td>4.9</td>
<td>6.4</td>
<td>8.2</td>
<td>10.9</td>
<td>39.3</td>
</tr>
<tr>
<td>4.62</td>
<td>4.29</td>
<td>5.3</td>
<td>6.3</td>
<td>8.2</td>
<td>9.0</td>
<td>11.5</td>
<td>37.9</td>
</tr>
<tr>
<td>4.29</td>
<td>4.03</td>
<td>7.1</td>
<td>8.5</td>
<td>10.1</td>
<td>11.5</td>
<td>15.0</td>
<td>43.7</td>
</tr>
<tr>
<td>4.03</td>
<td>3.83</td>
<td>7.7</td>
<td>9.7</td>
<td>11.7</td>
<td>13.6</td>
<td>17.7</td>
<td>46.8</td>
</tr>
<tr>
<td>3.83</td>
<td>3.66</td>
<td>6.6</td>
<td>9.0</td>
<td>10.8</td>
<td>13.1</td>
<td>18.6</td>
<td>52.1</td>
</tr>
<tr>
<td>3.66</td>
<td>3.52</td>
<td>6.1</td>
<td>8.8</td>
<td>12.3</td>
<td>15.0</td>
<td>22.3</td>
<td>54.5</td>
</tr>
<tr>
<td>3.52</td>
<td>3.40</td>
<td>7.6</td>
<td>11.3</td>
<td>14.8</td>
<td>18.9</td>
<td>27.1</td>
<td>63.9</td>
</tr>
<tr>
<td>3.40</td>
<td>3.30</td>
<td>7.7</td>
<td>12.5</td>
<td>17.9</td>
<td>22.6</td>
<td>31.3</td>
<td>66.2</td>
</tr>
<tr>
<td>3.30</td>
<td>3.20</td>
<td>8.9</td>
<td>15.8</td>
<td>21.4</td>
<td>28.0</td>
<td>40.5</td>
<td>75.1</td>
</tr>
<tr>
<td>3.20</td>
<td>3.12</td>
<td>10.5</td>
<td>17.6</td>
<td>25.3</td>
<td>32.2</td>
<td>44.7</td>
<td>79.4</td>
</tr>
<tr>
<td>3.12</td>
<td>3.04</td>
<td>12.5</td>
<td>20.6</td>
<td>28.6</td>
<td>35.7</td>
<td>51.0</td>
<td>83.2</td>
</tr>
<tr>
<td>3.04</td>
<td>2.97</td>
<td>15.7</td>
<td>28.0</td>
<td>36.0</td>
<td>43.6</td>
<td>57.7</td>
<td>86.1</td>
</tr>
<tr>
<td>2.97</td>
<td>2.91</td>
<td>17.1</td>
<td>28.3</td>
<td>38.4</td>
<td>47.3</td>
<td>62.9</td>
<td>93.0</td>
</tr>
<tr>
<td>2.91</td>
<td>2.85</td>
<td>20.8</td>
<td>32.7</td>
<td>43.1</td>
<td>54.4</td>
<td>68.7</td>
<td>93.6</td>
</tr>
<tr>
<td>2.85</td>
<td>2.80</td>
<td>19.5</td>
<td>33.0</td>
<td>46.2</td>
<td>56.0</td>
<td>70.8</td>
<td>95.6</td>
</tr>
<tr>
<td>2.80</td>
<td>2.75</td>
<td>21.3</td>
<td>37.5</td>
<td>49.5</td>
<td>58.4</td>
<td>73.6</td>
<td>96.4</td>
</tr>
<tr>
<td>2.75</td>
<td>2.70</td>
<td>23.5</td>
<td>40.1</td>
<td>52.6</td>
<td>62.5</td>
<td>77.3</td>
<td>97.3</td>
</tr>
<tr>
<td>All hkl</td>
<td>10.2</td>
<td>16.5</td>
<td>22.0</td>
<td>26.9</td>
<td>35.7</td>
<td>65.2</td>
<td>99.9</td>
</tr>
</tbody>
</table>

Table 6.2 The redundancies of each resolution shell.
b. Phenix and COOT

After a good data file (.sca) was generated from HKL2000, Phenix and COOT were used for structure determination and model building. Experimental parameters were specified as in Figure 6.4. The automatic structure determination with Autosol was successful. The final model had a Rwork of 0.2995, and Rfree of 0.3518. Out of the 600 residues, 395 were successfully built in according to the amino acid sequence I supplied. From the electron density map, it seemed that the built-in part has two domains which structurally resembled the head domain and body domain of the influenza A virus NP. The N-terminal residues 1-137 were not built in the final model, and there was a large region of extra density with no residues built in. The automatic model building was not

![Figure 6.4](image.png) The user-interface of Autosol and the parameters supplied.
successful for this region probably because there was no detected selenium sites within this density such that it was hard to determine the sequence registry. Another reason was that the density of the N-terminal domain was much weaker than the other parts of the protein, probably due to this region having a higher B-factor.

To manually build the residues into the N-terminal domain, I relied on the electron density map as well as the secondary structure prediction from the PREDICTPROTEIN online server (Figure 6.5 and 6.6). The

---

**Figure 6.5** Secondary structure prediction of the 1-135 residues on the N terminus. E in blue stands for residues making up a beta-sheet. H in red stands for residues making up a alpha-helix.

**Figure 6.6** Electron density map of the $\beta$-turn-$\beta$-turn-$\beta$ structure at the N-terminus of the ISAV-NP. A is with residues built-in and B is without.
model building was performed with the program COOT (Emsley and Cowtan, 2004). The secondary structure prediction could be trusted, because its prediction of the head-domain and body-domain agreed very well with the model that was automatically built. From the secondary structure prediction, I found that there is a β-turn-β-turn-β structure at the very N-terminus (Figure 6.5). From the density map, I found a corresponding β-turn-β-turn-β density (Figure 6.6). Because this was the only β-sheet structure that had not been built, I was certain about the correctness of the placement, except the sequence registry within the β-sheet structure, as well as the direction of the β-sheet. I noticed that there were two arginines in a row in the sequence of the N-terminal region. Arginine has a long side chain and it is often easily visualized in a density map. Therefore I tried to look for two residues in a row with long side chains. I did find two such residues in a row, and the registry in the β-sheet was determined (Figure 6.7). The direction of the β-sheet was determined by a trial-and-error process: one possibility was negated due to significant disagreement with the electron density and the other fit the density map well.

Figure 6.7 Electron density of Arg 20 and Arg 21 with residues built in. Those two were the first two that were built in to determine the registry of the sequence for the N-terminal domain.
Besides the N-terminal region, there were other regions I needed to build in manually, or make corrections where the residues were not so well built by the computer. Usually, after each round of building, Phenix.refine would be run to check if the $R_{\text{free}}$ improved. Phenix.refine also provides very helpful guidance for model building because it lists all the Ramachandran outliers, rotamer outlier, and C-beta outliers that need to be rebuilt or adjusted, based on the Molprobity web server. It took me about 50 iterations of manual model building and refinement, before obtaining a good final model.

The quality of the model was evaluated with the Phenix.validate program. The final 2.7Å model has an $R_{\text{work}}$ value of 0.2678, and an $R_{\text{free}}$ value of 0.2706. The RMS(angles) is 1.133, and the RMS(bonds) is 0.0091. The average B-factor is 59.10Å$^2$. The inherent structural flexibility of the N-terminal domain is confirmed by its high temperature factors, averaging 85.0Å$^2$ for the N-terminal domain, as compared to 54.8Å$^2$ for rest of the structure (Figure 6.8). The model has no Ramachandran outliers, and the percentage of Ramachandran favored residues is 95.6%.

---

**Figure 6.8** The B-factor plot generated by Phenix. The average B-factor of the N-terminal domain of ISAV-NP is much higher than the rest of the protein.
There are only 0.3% rotamer outliers, and no C-beta outliers. The structure coordinates has been deposited into the Protein Data Bank with a PDB ID 4EWC. All statistics during the data processing and structure determination are listed below.

<table>
<thead>
<tr>
<th>Table 6.4 Statistics of Data collection, phasing, and model building</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space group</strong></td>
</tr>
<tr>
<td><strong>Unit-cell dimensions</strong></td>
</tr>
<tr>
<td>a (Å)</td>
</tr>
<tr>
<td>b (Å)</td>
</tr>
<tr>
<td>c (Å)</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
</tr>
<tr>
<td>Last resolution shell (Å)</td>
</tr>
<tr>
<td>Rsym (%)</td>
</tr>
<tr>
<td>Average I / σI</td>
</tr>
<tr>
<td>Completeness (%)</td>
</tr>
<tr>
<td>Redundancy</td>
</tr>
<tr>
<td><strong>Phasing</strong></td>
</tr>
<tr>
<td># of sites</td>
</tr>
<tr>
<td>FOM</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
</tr>
<tr>
<td>Resolution</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt; / R&lt;sub&gt;free&lt;/sub&gt;</td>
</tr>
<tr>
<td># of R&lt;sub&gt;free&lt;/sub&gt; reflections</td>
</tr>
<tr>
<td># of protein atoms</td>
</tr>
<tr>
<td>Average B factor</td>
</tr>
<tr>
<td>R.m.s deviations</td>
</tr>
</tbody>
</table>
The crystal structure of the ISAV-NP

The final model of the ISAV-NP ΔC16 contains 495 amino acid residues in total, including residues 5-25, 27–38, 41-45, 61–75, 86–112, 133-182, 201-320, 325-497, 504-530, and 542–586. The structure of the nucleoprotein is mostly α-helical and each monomer has a curved, crescent-like shape (Figure 7.1, 7.2). The overall structure can be divided into 3 domains: an N-terminal domain, a head domain, and a body domain. The head domain and body domain are joined together by the polypeptide chain at three regions. The head and body domains are formed by non-contiguous polypeptide regions: the head domain is formed by residues 272–380 and 494–558, and the body domain consists of three polypeptide segments 113–271, 381–493, and 559–600. The N-terminal domain consists of amino acids 1 to 112. Although there is no detectable sequence similarity between the ISAV-NP and the influenza A NP, a structural comparison shows that the head and body domains of the ISAV-NP closely resemble those of the influenza A virus NP (Figure 7.3).

7.1 Dimerization via the tail loop

Although there is only one NP in each crystallographic asymmetric unit, the two NP molecules related by the crystallographic 2-fold symmetry axis form a biological dimer (Figure 7.1). The tail loop structure, which plays a critical role in influenza A virus NP oligomerization, is also found in the ISAV-NP, despite the two NP proteins being
Figure 7.1 ISAV-NP crystal structure. Three orthogonal views (A-C) of an NP dimer. One subunit is colored in green, whereas the other subunit is colored differently for each domain with blue for the N-terminal domain, yellow for the head, red for the body, and magenta for the tail loop. In (C), the magenta tail loop is shown with a space-filling model to highlight the extensive interaction mediated by the tail loop.
Figure 7.2 ISAV-NP monomer. The molecule is colored continuously from blue to red from the N-terminus to C-terminus, respectively. Secondary structural elements are numbered.
Figure 7.3 ISAV-NP (red) superimposed onto the influenza A virus NP (cyan, PDB ID: 2IQH). The head and body domains of the ISAV-NP highly resemble those of the influenza A NP.
crystallized in different oligomerization states (dimer vs. trimer). Formed by residues 504–530 at the back of the molecule, the tail loop of the ISAV-NP folds into an extended-hairpin that protrudes into a deep pocket in the neighboring molecule to mediate the dimer formation. The interaction made by the tail loop is primarily hydrophobic in nature. Near the tip of the tail loop, the only charged residue Asp513 makes hydrogen bonds to four main-chain nitrogen atoms in the opposing peptide chain from the same tail loop, thus stabilizing the formation of the hairpin. Deletion of the ISAV tail loop (Δloop or Δ504-530) resulted in the formation of NP monomers exclusively (Figure 7.4), suggesting that such a tail loop structure is important for NP oligomerization for myxoviruses in general.

7.2 The RNA-binding groove

Analogous to the influenza A virus NP, there is a potential RNA-binding cleft between the head and body domains of the ISAV-NP. The groove is positively charged according to its calculated electrostatic potential, and is covered with a large number of basic amino acid residues, which presumably interact with the negatively charged phosphor-diester backbone of bound RNA (Figure 7.5). All of these basic residues, except those from the N-terminal domain, are conserved in the influenza A virus NP (Figure 7.6). The aromatic residue F274, analogous to Y148 from the influenza A virus NP (Ye et al., 2006), is found at one end of the RNA-binding cleft, and may mediate nucleotide base stacking.
**Figure 7.4** Gel filtration chromatograms for the ISAV-NP proteins. The eluted positions of three protein standards are marked by arrows. According to the elution volumes, the full length ISAV-NP and the ISAV-NP ΔC16 exist as only dimers, whereas ISAV-NP ΔLOOP exists exclusively as monomers.
Figure 7.5 Calculated electronic potential for an NP monomer. Positively charged residues are highlighted.
Figure 7.6 ISAV-NP secondary structure assignment. The α-helices are shown by cylinders and the β-strands are represented by arrows. Conserved charged residues from the RNA binding groove are highlighted in blue. The conserved aromatic residue F274 in the groove is shown in red. The NP sequences from ISAV and the influenza A virus were manually aligned, based on tertiary structures.
To test whether the ISAV-NP binds RNA in the same manner as the influenza A virus NP, we made two double mutants: K185A/R186A and K296A/R299A. These two pairs of basic residues are from two flexible loops in the potential RNA binding groove that are partially disordered in the ISAV-NP structure. At equivalent positions in the influenza A virus NP, R74/R75 and R174/R715 were found to play critical roles in RNA binding (Ng et al., 2008). As expected, mutations to these residues resulted in a dramatic decrease in RNA binding affinity, as the $K_d$ changed from 24.5nM for the wild-type (wt) protein to 355 nM and 472 nM for the K185A/R186A and K296A/R299A mutants, respectively. These results indicate that NPs from both ISAV and the influenza A virus likely bind RNA with their positively charged groove.

7.3 The N-terminal domain

The N-terminal domain of the ISAV-NP is a unique structural feature that was not observed in the crystal structures of either the influenza A or the influenza B virus NP. Consisting of three anti-parallel $\beta$-strands and two $\alpha$-helices, the ISAV N-terminal domain adopts an extended shape and interacts with the rest of the structure mainly through the long $\alpha$-helix made of residues 86-112 (Figure 7.2). This large $\alpha$-helix partially covers the potential RNA binding cleft in between the head and body domain. Deleting the entire N-terminal domain had little effect on the equilibrium-state of the RNA binding, as the $\text{wt}$ and $\Delta N111$ NP exhibited very similar RNA binding affinities ($K_d = 24.5$ nM for $\text{wt}$ vs. 33.6 nM for $\Delta N111$, using a 24-nt RNA) (see next chapter for more
details). Nevertheless, it is plausible that the N-terminal domain may regulate the
dynamics of the NP RNA binding. For instance, it may help to attract RNA substrates
and/or serve as a cap to prevent the dissociation of RNA prior to being locked in a
bound position. We speculate that the long α-helix from the N-terminal domain may
need to be displaced from its current location upon RNA binding. The inherent
structural flexibility of the N-terminal domain is evident from its high temperature
factors, averaging 85.0 Å² for the N-terminal domain, as compared to 54.8 Å² for the rest
of the structure (Figure 6.13).

Multiple-sequence alignments show that an N-terminal domain is consistently found in
every orthomyxovirus NP, but it is highly variable in both size and sequence (e.g. 21aa for
the influenza A virus, 72aa for the influenza B virus, and 111aa for ISAV). Furthermore,
the N-terminal domain is structurally disordered in both the influenza A and influenza B
virus NPs (Ng et al., 2012b; Ng et al., 2008; Ye et al., 2006). In the influenza A virus, the
first 20 amino acids contain a phosphorylation site and a nuclear localization signal (NLS)
that are important for the intracellular trafficking of the NP (Arrese and Portela, 1996;
Wang et al., 1997). Such functions have not been attributed to the ISAV N-terminal
domain. Although two mono-partite nuclear localization signals 230RPKR233 and
473KPKK476 have been identified in the ISAV-NP (Aspehaug et al., 2004), both were
mapped to the surface of the body domain. The location of the N-terminal domain
suggests that it would be situated on the exterior of assembled RNPs, and therefore may
help to mediate interactions with the viral matrix protein and/or NS2 to facilitate the
intracellular trafficking of viral RNPs.

7.4 The missing density of nucleic acids

Although I tried to purify and crystallize the complex of the ISAV ΔC16 and poly-dC18 DNA, no density for the DNA can be seen in the final electron density map. It is very likely that the complex of NP:DNA has been successfully produced and purified, based on the facts that the molecular weight of the final product increased as shown by a shifted peak to the left in the size-exclusion chromatogram, and that the ratio of the UV absorbance 260/280 of the purified product showed that DNA was incorporated. Regarding the missing DNA density, there could be many possible explanations.

First, it is likely that the DNA was not present in the final crystal, because the high-salt of the crystallization condition (0.8-1.0 M) might have caused dissociation of the initially bound DNA during the crystallization process. At least 0.8M of lithium chloride was required for the crystallization of well-diffracting crystals. The apo form of the ISAV ΔC16 in the absence of DNA was resistant to crystallization under similar condition. Therefore, DNA was required for the crystallization in a role like “catalyst”, and the DNA was somehow finally dissociated from the protein in the process of crystallization or during crystal growth.

Another possibility was that the DNA was still in the crystal, but could not be “seen” in
the final electron density map. It might be possible that the DNA density was “averaged out” in the final electron density map due to some non-specific binding. Or, instead of non-specific binding, it is also possible that the occupancy of DNA was low. The small amount of DNA was still able to increase the average molecular weight shown in the size-exclusion chromatography, and make the UV absorbance 260/280 ratio more “complex-like”. However, the low occupancy of DNA reduced the relative intensity of the electron density, making it less-likely to be identifiable in the final density map.

To investigate which explanation is more likely, I performed the following experiment. I harvested 20 crystals larger than 0.1mm in size. I washed all of them by transferring them into a wash buffer (0.8M LiCl, Citrate 4.0, 25% PEG6000) to get rid of at residual protein or DNA on the surface. The wash was performed five times. Then the crystals were transferred to high-salt buffer (2M NaCl, 0.1M HEPES 7.5, 10% glycerol (v/v)) to dissolve them. This solution with the dissolved crystals was measure with a UV spectrometer to determine the UV absorbance at 260/280. The results showed that the ratio of UV280 to UV260 was close to 1.8, suggesting that DNA was not present in the final crystals.
7.5  ISAV-NP is a good model to study the RNA-binding

As mentioned before, the structure of the ISAV-NP head domain and body domain highly resembles the structure of the influenza A virus NP. The ISAV-NP always exists as homogenous dimers in solution, which is advantageous in not only the crystallization of NP-RNA complex, but also the biophysical characterization of the NP-RNA binding affinity and stoichiometry. The ISAV-NP dimer is stable in solution, and it is relatively resistant to protease digestion. In addition, unlike the influenza A virus NP, the ISAV-NP does not aggregate upon being treated with nucleic acids. Finally, the ISAV-NP is easy to express using *E. coli* with a great yield rate, which provides sufficient material for various experiments. Therefore, I propose that the ISAV-NP is an excellent model to study the influenza A NP, especially the NP:RNA complex structure and the interaction between the NP and RNA.
8. Characterization of the RNA binding of the ISAV-NP

Unlike the influenza A virus NP, which undergoes dynamic equilibrium between oligomers and monomers (Ye et al., 2006), the ISAV-NP dimer is stable and does not dissociate under a wide range of conditions when tested in vitro. Furthermore, the ISAV-NP forms homogenous dimers, unlike the influenza A and B virus NPs which form multiple oligomeric species in solution (Ng et al., 2012b; Ng et al., 2008; Ye et al., 2006). These unusual features make the ISAV-NP an excellent model system for studying the NP RNA binding affinity as well as stoichiometry; two parameters that are crucial for our understanding of orthomyxovirus RNP structure and assembly. In this thesis project, fluorescence polarization assays (FP) were used to determine the affinity and stoichiometry of NP RNA binding with synthetic RNA oligos. Electron microscopy (EM) and dynamic light scattering (DLS) experiments were employed to confirm our results.

8.1 Fluorescence Polarization

I first measured the RNA-binding affinity of three NP variants, the wt NP, NP ΔN111, and NP ΔC16. ΔN111 was chosen because it lacks the N-terminal domain, and thus its structure is comparable to that of the influenza A virus NP. Because genome-wide stoichiometry estimation revealed that each influenza A virus NP binds to 24 nts of RNA (Ortega et al., 2000), the oligonucleotide (AC)\textsubscript{12} (consisting of 12 AC repeats) was
chosen for our initial assays. For all proteins, the binding affinity increased, shown by a decreased $K_d$, as the length of the RNAs increased (Figure 8.1A-C).

Interestingly, the wt NP and ΔN111 exhibited similar RNA-binding affinities (Figure 8.1A, B), suggesting that the N-terminal domain is not actively involved in equilibrium-state RNA binding. However, it is plausible that the N-terminal domain may regulate the dynamics of the NP RNA binding. For example, it may function in attracting RNA substrates and/or serving as a cap to prevent the dissociation of RNA prior to being locked in a bound position. I speculate that the long $\alpha$-helix (aa86-112) from the N-terminal domain may need to be displaced from its current location upon RNA binding. This speculation is supported by the high B-factors of the N-terminal domain, averaging 85.0 Å$^2$ for the N-terminal domain, as compared to 54.8 Å$^2$ for rest of the structure (Figure 6.8).

On the other hand, NP ΔC16 exhibited a much stronger RNA binding affinity than both the wt NP and ΔN111 (Figure 8.1), indicating that the acidic C-terminal tail negatively regulates NP-RNA binding. It is possible that this highly-negatively charged C-terminal tail interacts with the RNA-binding groove which is rich in basic residues (Arg, Lys) and thus positively charged. In other words, the C terminal tail may inhibit the protein-RNA binding in a competitive manner. An acidic C-terminal tail is also conserved in the Influenza A NP and Influenza B NP, so it is likely to play an important biological role, and this competitive inhibition may be universal in orthomyxoviruses.
Figure 8.1 ISAV-NP RNA binding. (A-C) RNA binding affinity measurements for the wt NP (A), ΔN111 (B) and ΔC16 (C). FA was performed using three RNA oligos each containing 20, 24, and 28 nts. The binding curves are plotted using the same protein concentration range. For the ΔC16 mutant, an inset is added in (C) to provide a more spread-out view at lower protein concentrations.
When the RNA oligos were shortened from 24 nts to 20 nts, the binding affinity of the NP dropped sharply with a ~10-fold increase in $K_d$ (Figure 8.1A). However, when the size of the RNA oligo was increased from 24nt to 28nt; there was a smaller change in the NP-RNA binding affinity (Figure 8.1A). Similar results were observed for the ΔN111 and ΔC16 mutants: the dissociation constants between the protein and the 24nt-RNA were significantly affected when the RNAs became shorter but not when the RNAs became longer (Figure 8.1B, C). Therefore, the 24nt-long RNA has saturated the NP-RNA binding site, and the small increase in binding affinity from 24-nt to 28-nt RNA was probably due to non-specific binding.

To determine whether each 24-nt RNA binds to a single NP subunit or to both subunits in a NP dimer, I used the same fluorescence polarization assay to quantify binding stoichiometry. To ensure stoichiometric binding, I adjusted the experimental conditions such that the total concentration of RNA was significantly above the binding dissociation constant $K_d$. When the wt NP was titrated in the solution, the fraction of bound RNA, as noted from the millipolarization (mP) signal, first increased linearly and then gradually approached saturation (Figure 8.2A). Using either the inflection point extrapolation or direct fitting of the general titration equation, I arrived at the same conclusion that each 24-nt RNA binds to an NP dimer. Subsequently, a similar experiment using a 48-nt RNA gave a stoichiometry of two NP dimers per 48-nt long RNA (Figure 8.2B), consistent with our earlier result of each NP monomer binding to ~12 nts of RNA.
To rule out the possibility that one NP subunit binds 24 nts of RNA and that RNA binding by one NP subunit allosterically inhibits RNA binding by the other subunit in the NP dimer, a monomeric ΔLOOP mutant was used for RNA binding assays as described above (Figure 8.3). Circular dichroism spectra indicated that the overall secondary structure content of the ΔLOOP mutant was indistinguishable from that of the wt NP dimer, suggesting that the overall tertiary structure is preserved (Figure 8.4). As expected, the RNA binding affinity of the ΔLOOP monomer was very similar to that of the wt NP for a 12-nt long RNA, which supposedly contains only one NP binding motif \( K_d = 488.4 \text{ nM for } \text{wt vs. } 1076.1 \text{ nM for } \Delta\text{LOOP NP} \). However, the protein binding stoichiometries of the 12-nt and 24-nt RNAs were unexpectedly different. While each 12-nt RNA bound to one ΔLOOP NP monomer, each 24-nt RNA recruited two NP ΔLOOP monomers when fully saturated (Figure 8.3). Therefore, results obtained from the ΔLOOP mutant are consistent with our earlier conclusion based on the wt NP dimer.
**Figure 8.2** RNA binding stoichiometry measurements for the wt NP. The calculated ratios by curved fitting were shown in boxes. NP:RNA complexes assembled in each experiments are schematically shown with blue handcuffs for the NP dimers and grey curves for RNA molecules.
**Figure 8.3** RNA binding stoichiometry measurements for the ΔLOOP monomeric mutant. The calculated ratios by curved fitting were shown in boxes. NP:RNA complexes assembled in each experiments are schematically shown with blue handcuffs for NP dimers and grey curves for RNA molecules.
Figure 8.4 Far-UV CD Spectra of the ISAV-NP (blue) and its two mutants, ΔLOOP (red) and ΔC16 (green). The secondary structures of the three proteins (all at 5 μM) were probed in the 190–250 nm range. The protein buffer was 200 mM potassium phosphate, pH 7.5. Measurements were done at room temperature.
8.2 Electron Microscopy

Negative-staining electron microscopy (EM) was performed to confirm the RNA-binding stoichiometry by measuring the number of NP molecules associated with each RNA molecule. As shown, the ISAV-NP formed dimers in the absence of RNA, with each NP molecule assuming the shape of a 4x5nm oval (Figure 8.5 E, F). When the 24-nt RNA was added to the NP at the 1:2 molar ratio (i.e. one RNA per NP dimer), the same dimeric assemblies were observed (Figure 8.5 C, D). However, when the 48nt-RNA were added to the ISAV-NP at the 1:4 ratio (i.e. one RNA for four NP subunits), we observed mostly tetramers, which were presumably formed by the simultaneous binding of two NP dimers to one 48-nt RNA (Figure 8.5 A, B).
Figure 8.5 ISAV-NP:RNA complexes characterized by electron microscopy. (A-F) EM images of the NP bound to a 48-nt RNA (A, B), NP bound to a 24-nt RNA (C, D), and free NP (E, F).
8.3 Dynamic Light Scattering

Similar results of the NP-RNA binding were also obtained from size measurements using dynamic light scattering (DLS) (Figure 8.6). Assuming that the ISAV-NP adopts a roughly globular shape, the molecular weight of the wt NP dimer was 150.9 ± 13.0 kD by DLS measurement. The calculated molecular weight of a dimer is 136.5 kD, and the slight difference could be explained by the non-globular shape of the dimer. When the 48-nt RNA was added to the NP sample, the molecular weight of the complex was increased to 299.7 ± 16.2 kD, which matches closely to the calculated molecular weight of two ISAV-NP dimers binding to a 48-nt poly(C) RNA (i.e., \(2 \times 136.5 + 14.8 = 287.8\) kD). The results by EM and DLS once again confirm that one ISAV-NP molecule binds to a 12-nt RNA.

![DLS](image)

*Figure 8.6* Molecular weights of NP and NP:RNA complexes as determined by DLS, the average molecular weight with standard deviation (\(N = 20\)) is labeled on the top of each column. Molecular interpretations were made for each figure with blue handcuffs for the NP dimers and grey curves for the RNA molecules.
8.4 Functional implication

Since only the head and body domains of the ISAV-NP affect equilibrium-state RNA binding, the ISAV-NP serves as a valid model for studying the NP-RNA binding of orthomyxoviruses. We conclude that each ISAV-NP monomer, and likely the influenza A virus NP, bind to ~12 nts of RNA. This stoichiometry of binding is also supported by the observation that the potential RNA-binding groove of the ISAV and the influenza A virus NP is ~60 Å long. Assuming that bound nucleotides are ~5Å apart, each NP molecule should bind to no more than ~12 nts of RNA, as suggested by our stoichiometric measurements. Indeed, RNase digestion of an influenza A virus mini-RNP produced RNA fragments that were ~18-nt in size (Neumann and Hobom, 1995), indicating that the size of the RNA protected by each NP should be shorter than 18 nts long.

Considering that on average there are ~24 nts of RNA for each NP, and that each NP molecule only binds to 12 nts of RNA, a significant amount of the orthomyxovirus genomic RNAs should be NP-free. Such NP-free RNA regions are probably rich in secondary structure, so binding cannot occur, because NPs only bind to single-stranded RNA with high affinity (Baudin et al., 1994). Thus, we envision that there is an uneven distribution of NP along the viral RNAs that varies depending on the secondary structure of the RNA. Because nascent v/c RNA molecules are encapsidated by NP co-transcriptionally, we assume that only local secondary structures occur in RNP.

These RNA structures likely play a critical role in the packaging of the viral genome.
during the assembly process, possibly by forming specific higher-order structures. These structures would allow the eight different RNPs to recognize each other and form the “7+1” array previously seen in EM images of budding influenza A virus particles (Harris et al., 2006; Noda et al., 2006; Yamaguchi et al., 2008). Thus, the 2nm thick filamentous masses connecting neighboring RNPs in budding influenza A viruses are likely RNA structures formed by packaging signals. Because the RNA packaging signals are usually found near both ends of the vRNA molecules (Hutchinson et al., 2010), each RNP may interact with at least two neighboring RNPs to build up a network of interactions that ensure specific genome packaging.

The stoichiometry of one NP per 12 nts of RNA is different than those reported for rhabdovirus and paramyxovirus, in which each N protein binds 9 and 7 nts of RNA, respectively (Albertini et al., 2006; Green et al., 2006; Tawar et al., 2009). It is also worth mentioning that the RNPs of the Mononegavirales are strictly stoichiometric complexes, with N proteins distributed evenly along the N-RNA complexes; whereby vRNA is completely unwound and protected by the NP (Tawar et al., 2009). Indeed, the RNPs of the Mononegavirales are highly resistant to RNase treatment, but viral RNA in the orthomyxovirus RNPs are sensitive to RNases (Baudin et al., 1994; Hsu et al., 1987). Our study thus establishes that orthomyxovirus RNPs are unique genomic protein-RNA complexes, which are enriched in free RNAs, a feature optimally suited for multi-segment RNA packaging by allowing direct RNA-RNA interactions.
9. Major conclusions

The crystal structure of the ISAV-NP is determined to 2.7Å resolution using single-wavelength anomalous diffraction (SAD) and SeMet-derivatived protein. The overall structure can be divided into three domains: a head domain, a body domain, and an N-terminal domain. Structural comparison shows that the head and body domains of the ISAV-NP closely resemble those of the influenza A virus NP. Similar to the influenza A virus NP, the oligomerization of the ISAV-NP is also mediated by a “tail loop” structure which inserts into a deep pocket in the neighboring molecule. Deletion of the ISAV tail loop (Δloop or Δ504-530) results in the formation of NP monomers exclusively, suggesting that such a tail loop structure is important for NP oligomerization of myxoviruses in general.

The N-terminal domain of the ISAV-NP is a unique structural feature that was not observed in the crystals structures of either the influenza A or influenza B virus NP, although a multi-sequence alignment shows that an N-terminal domain with variable size and sequence is consistently found in every orthomyxovirus NP. RNA-binding studies show that the N-terminal domain had little effect on the equilibrium-state RNA binding. A B-factor plot shows that the N-terminal domain is much more flexible than the other part of the protein. Therefore, it is plausible that the N-terminal domain may regulate the dynamics of the NP RNA binding by attracting RNA substrates and/or serving as a cap to prevent the dissociation of RNA prior to being locked in a bound position.
Similar to the influenza A virus NP, there is a potential RNA-binding groove in between the head and body domains of the ISAV-NP. The groove is rich in basic amino acid residues, and all of these basic residues are conserved in the influenza A virus NP. The RNA-binding affinity of two ISAV-NP double mutants, K185A/R186A and K296A/R299A, have been tested, because these mutations affected residues analogous to the influenza A virus NP residues R74/R75 and R174/R175 that have been found to be critical for RNA binding. As expected, mutations to these residues result in a dramatic decrease (>10-fold) in the RNA binding affinity, showing that the NPs from both ISAV and the influenza A virus likely bind RNA in the same fashion.

The structure of the ISAV-NP head domain and body domain highly resembles the structure of the influenza A virus NP. The ISAV-NP always exists as homogenous dimers in solution, which allows not only the crystallization of the NP-RNA complex, but also the biophysical characterization of the NP-RNA binding affinity and stoichiometry. Fluorescence polarization assays (FP) have been used to determine the affinity and stoichiometry of the NP RNA binding with synthetic RNA oligos. Surprisingly, the RNA binding analysis reveals that each NP binds ~12 nts of RNA, shorter than the 24-28 nts that was originally estimated for the influenza A virus NP. The 12-nt stoichiometry has been further confirmed by results from electron microscopy and dynamic light scattering.

The RNA-binding assays have proven that only the head and body domains of the
ISAV-NP affect equilibrium-state RNA binding, so the ISAV-NP serves as a valid general model for studying the NP-RNA binding of orthomyxoviruses. We conclude that each ISAV-NP monomer, and likely the influenza A virus NP, binds to \(~12\) nts of RNA. Considering that on average there are \(~24\) nts of RNA for each NP, and that each NP molecule only binds to \(12\) nts of RNA, a significant amount of the orthomyxovirus genomic RNAs should be NP-free. Such NP-free RNA regions are probably rich in secondary structures which likely play a critical role in the packaging of viral genome during the assembly process, possibly by forming specific higher-order structures.

This study thus establishes that the orthomyxovirus RNPs are unique genomic protein-RNA complexes, which are enriched in free RNAs, a feature optimally suited for multi-segment RNA packaging by allowing direct RNA-RNA interactions. In addition, results from this thesis project show that the ISAV-NP is an excellent model for studying the influenza A NP, especially the NP:RNA complex structure and the interaction between the NP and RNA.
10. Future directions

10.1 Structure determination of the ISAV-NP-RNA complex

Because we are fascinated about using the ISAV-NP as an ideal model to study the influenza A virus NP, I plan to keep pursuing the structure determination of the ISAV-NP:RNA complex. Because previous attempts to crystallize a purified complex only gave crystals of the protein without nucleic acids, stronger interaction between the protein and RNA is needed. To increase the RNA-binding affinity, one can increase the RNA length, reduce the salt concentration in the buffer, or cross-link the RNA to the protein. Using a longer RNA may not be a top priority since it may introduce undesirable flexibility. Although lowering the salt concentration can significantly increase the binding affinity, it also affects the protein solubility and thus protein may precipitate out of solution over time, because the salt concentration of the buffer is at its lowest possible level. Therefore, I plan to crystal the NP:RNA complex by RNA cross-linking.

It has been reported several time that complexes of disulfide cross-linked protein and nucleic acid led to a successful crystallizations and structure determination (Corn and Berger, 2007; Ho et al., 2006; Huang et al., 1998; Qi et al., 2009; Verdine and Norman, 2003; Yang et al., 2011). During the cross-linking experiments, DNA/RNA can be “trapped” on a protein by using a disulfide cross-linking strategy, which is a structurally passive tool that acts only to reduce the chance of dissociation, because the cross-linking itself seems to be mobile in conformation, and the nearby regions are not distorted (He
I will try to follow this cross-linking protocol and apply it to the crystallization of the ISAV-NP:RNA complex. The RNA used will have a modified cytosine base with a thiol tether introduced at the N4 position. This will allow disulfide cross-linking between the thiol tethered RNA and a surface cysteine residue from the ISAV-NP. One amino acid residue at the predicted RNA-binding groove would be mutated to cysteine, while all the other 4 surface cysteines will be mutated to serines/alanine to prevent unwanted cross-linking. The ISAV-NP cross-linked with RNA will be separated from the unlinked protein with an ion-exchange column using the FPLC. The purified cross-linked NP:RNA complex will then be used for crystallization screens. If quality crystals are obtained, the structure of the NP:RNA complex can be determined by molecular replacement. Once determined, the complex structure would provide invaluable insights into understanding the functions of the ISAV-NP as well as the influenza A virus NP, as well as the design and development of anti-influenza drugs by disrupting the NP-RNA interaction.
10.2 Structure determination of the ISAV-NP monomer

The NP can be a good target for drugs designed against influenza viruses, because its sequence is one of the most conserved among all the influenza proteins. NP oligomerization is important for the influenza virus because it provides a major structural framework for RNP assembly. Drug compounds that promote aberrant NP aggregation can effectively inhibit influenza A virus replication in cell cultures, suggesting that the NP is a valid target for anti-influenza therapy (Gerritz et al., 2011; Kao et al., 2010).

To understand the mechanism by which the NP oligomerizes and how NP prevent self-aggregation prior to RNP encapsidation, it is important to have the structures of oligomerized NP and monomeric NP. A monomeric mutant with a single amino acid substitution (i.e. R416A) in the tail loop of the influenza A virus NP was found to undergo some structural rearrangements in the C-terminal part of the polypeptide, but the structural configuration around the potential RNA binding groove remains unaltered (Chenavas et al., 2013). Although we have shown that the ISAV-NP monomer and dimer have similar secondary structure by circular dichroism and that they bind to RNAs similarly by fluorescence polarization, it is still possible that the NP undergoes confirmation changes from “self-inhibitory” state to the oligomerized state.

The ISAV Δloop mutant proteins exists as NP monomers exclusively, but it may not be a good target to study the monomer structure because no crystallization conditions has ever been obtained from the screening of thousands of conditions. It seems that the
presence of the tail loop may be necessary for the successful crystallization of the
ISAV-NP monomer. Therefore I am looking for mutagenesis strategies to prevent the
ISAV-NP from oligomerizing. The point mutation R416A of the influenza A virus NP
totally abolished oligomerization (Ye et al., 2006), suggesting that it is possible to obtain
only ISAV-NP monomers by mutating key residues on the tail loop. One good candidate
is Asp513, the only charged residue close to the tip of the tail loop, which makes
hydrogen bonds with four main-chain nitrogen atoms in the opposing peptide chain
from the same tail loop to stabilize the formation of the hairpin.

Ultimately, a comparison of the monomer structure and dimer structure would provide
extraordinary insights into the biological processes of NP oligomerization, genome
replication, as well as RNP assembly. Also, structure-based drug design can be employed
to design small molecules that disrupt the NP oligomerization, which could be an
effective way to fight against the influenza viruses.
Supplementary information

S1. Structural studies on other ISAV proteins

In Dr. Tao’s lab, I have also attempted to study the structure of other ISAV proteins, including the M1, M2, NS1, and NEP. All of these proteins were found to have similar counterpart in the influenza virus, and therefore studies on these ISAV proteins can help us to understand how these proteins function as well as provide useful hint in anti-flu drug design or vaccine development.

S1.1 ISAV-M1

The M1 protein is a multi-functional protein. It plays significant roles in virus replication, from virus entry to assembly and the budding of the virus particle (Cottet et al., 2011; Palese and Shaw, 2007). M1 binds to RNPs by interacting with the NP and linking it to the virus envelop. Also, the NEP (Nuclear Export Protein, also call NS2) is thought to be important to promote the nuclear export of the RNP complex, which is targeted to the virion assembly site (Boulo et al., 2007; Portela

Figure S1.1  Size-exclusion chromatography showing the interaction between M1 and mono-dispersity and SDS-PAGE showing the purity of M1.
and Digard, 2002). To date, only the structure of the N-terminal portion of the influenza A virus M1 is available (Sha and Luo, 1997). Based on the high similarity between the ISAV and the influenza A virus, it is possible that the ISAV-M1 structure resembles the influenza M1 protein. Therefore I attempted to express, purify, and crystallize the ISAV-M1 protein.

The M1 gene was cloned from the cDNA of the 8th ISAV gene segment and inserted into a pET28b+ vector (Novagen). The ISAV-M1 protein was expressed in the Rosetta 2 cell (Novagen). The protein was purified with Ni-NTA columns, HP SP columns, and size-exclusion columns. The purity was good and the sample was mono-dispersed (Figure S1.1). The ISAV-M1 protein has been successfully crystallized (Figure S1.2) and the diffraction of the resulting crystal has been tested. The M1 crystal diffracted to 3Å at our home source.

SeMet-derivative M1 protein was prepared in the same way as the ISAV NP, for phasing with Single-wavelength Anomalous Dispersion. The SeMet M1 was crystallized in 0.2M MgCl2, 0.1M HEPES 7.5, and 20% PEG3350 (v/v) at

**Figure S1.2** Crystals of the ISAV-M1. Single crystals can be separated from the cluster, for X-ray diffraction experiments.
4°C. Full size crystals were observed after two weeks. Crystals were clustered, but single crystals can be separated from the cluster, for X-ray diffraction experiments.

X-ray diffraction was performed at a Synchrotron source (Advanced Photon Source, Argonne National Laboratory, beamline 21-ID-F) at a wavelength of 0.97872Å (Figure S1.3). Data was collected with an oscillation angle of 1 degree for a 360-degree collection. The data was processed with HKL2000, and the statistics is shown in Table S1.1.

**Figure S1.3** X-ray diffraction of SeMet-M1 crystal at Synchrotron source (APS, 21-ID-F).

<table>
<thead>
<tr>
<th>Data collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
</tr>
<tr>
<td>Unit-cell dimensions</td>
</tr>
<tr>
<td>a (Å)</td>
</tr>
<tr>
<td>b (Å)</td>
</tr>
<tr>
<td>c (Å)</td>
</tr>
<tr>
<td>α (°)</td>
</tr>
<tr>
<td>β (°)</td>
</tr>
<tr>
<td>γ (°)</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
</tr>
<tr>
<td>Last resolution shell (Å)</td>
</tr>
<tr>
<td>Rsym (%)</td>
</tr>
<tr>
<td>Average I / σI</td>
</tr>
<tr>
<td>Completeness (%)</td>
</tr>
<tr>
<td>Redundancy</td>
</tr>
</tbody>
</table>

**Table S1.1** Statistics of data collection of ISAV SeMet-M1 crystal
The scaled data was analyzed by Phenix.xtriage, which suggested that there was very strong pseudo translational symmetry (PTS) which impeded successful detection of the heavy atoms (Se) and thus no correct solution could be found. PTS means two or more non-crystallographic symmetry-related molecules are in similar orientations that are indistinguishable by the program. The Patterson function was usually used to show the

<table>
<thead>
<tr>
<th></th>
<th>X-fraction</th>
<th>Y-fraction</th>
<th>Z-fraction</th>
<th>Peak height</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak1</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>136.36</td>
</tr>
<tr>
<td>Peak2</td>
<td>0.5000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>34.20</td>
</tr>
<tr>
<td>Peak3</td>
<td>0.4609</td>
<td>0.0000</td>
<td>0.7592</td>
<td>16.76</td>
</tr>
<tr>
<td>Peak4</td>
<td>0.0374</td>
<td>0.0000</td>
<td>0.2306</td>
<td>15.70</td>
</tr>
</tbody>
</table>

Table S1.2 Peaks from Patterson map of M1 data with a cut off of a height of 15Å. Other than the peak at origin, there are 3 significant “off-origin” peaks, showing strong PTS along both the X-axis and Z-axis. “non-origin” peaks which correspond to the relative displacement of these molecules. A patterns map was calculated with the CCP4 Program Suite, and the peaks were listed in Table S1.2.

The peaks were listed with the corresponding information of their height and fractional distance from the origin along each axis. Peak1 is the origin peak, and there are 3 significant “off-origin” peaks, showing strong PTS along both the X-axis and Z-axis. Peak2 was a “perfect” off-origin peak with a fractional distance of exactly 0.5 along the X-axis, showing that there were two M1 molecules along the X-axis assuming the same orientation within each unit cell. From peak3 and peak4, the positions were at 1/4 and
3/4 on the Z-axis, suggesting there might be 4 molecules along the Z-axis. According to the Matthew’s coefficient, it was most likely that there were 8 molecules in each unit cell, which was consistent with the inference from the PTS information (i.e., $2 \times 4 = 8$). It was unexpected that no significant peak was found at the 1/2 position along the Z-axis, which might be explained by the packing along the Z-axis being less ordered at the imperfect fractional distances along the X-axis of peak3 and peak4.

To get around the PTS problem, I have tried to re-index the data with a small unit cell to avoid the undesirable translation within the large unit cell. However, none of the re-indexed data showed any improvement no matter how I tried to re-index along the X-axis, Z-axis, or both. I have tried to solve the structure by molecular replacement with the influenza M1 at the search model, but it was not successful either. I have also tried to use the MR-Rosetta program which could extend the range of molecular replacement from the structure-modeling tools, but the PTS problem seemed to prevent successful structure determination, still. Therefore, I am now searching for new conditions to crystallize the M1 protein, hoping that the space group would be different and that the PTS would not exist.
S1.2 ISAV-M2

The ISAV-M2 is a protein translated from an alternatively sliced mRNA which is also transcribed from gene segment 8. The M2 protein is thought to be the membrane protein tetramerized to form the ion channel in the viral membrane. The M2 protein of the influenza A virus is a pH-controlled proton channel, which is activated by endosomal acidification to inject protons into the virus interior and lead to the dissociation of M1 from the vRNP. The structure of the influenza M2 protein has been determined using X-ray crystallography and NMR (Acharya et al., 2010; Hu et al., 2010; Schnell and Chou, 2008; Sharma et al., 2010), but it would still be interesting to see if the ISAV-M2 structure shares any homology with the structure of the influenza M2.

Figure S1.4 Purification of the ISAV-M2. A. Fractions of ISAV-M2 eluted from Ni-NTA column, showing a decent expression level and purity. B. Chromatography curve of ISAV-M2 from the size-exclusion column.
Similar to the ISAV-M1, the M2 gene was inserted into pET28b+ vector (Novagen), and the ISAV-M2 protein was successfully expressed in the Rosetta 2 cell (Novagen). The ISAV-M2 protein was purified with Ni-NTA resin, HP SP column and size exclusion column. The elusion volume of the ISAV-M2 was around 70ml, indicating that ISAV-M2 formers a tetramer, which was consistent with the tetramerization of the M2 protein in the influenza A virus. It was also noticed that the right shoulder of the ISAV-M2 peak was asymmetric when compared to the left shoulder. That “flat tail” on the right was probably caused by some mis-folded protein, since the ISAV-M2, as a membrane protein, may need some detergent to stabilize and keep it well-folded. Large-scale screens for crystallization conditions have been performed, but no crystallization has been identified. A detergent screen should be performed considering that M2, as a membrane protein, might need to be stabilized with detergents.
The ISAV-NS1 protein is encoded by gene segment 7. The NS1 protein functions to inhibit the splicing of host pre-mRNA, so as to ensure the production of viral proteins. The structure of the full length influenza NS1 has been determined by X-ray crystallography, which showed a dimerized NS1 with a dsRNA-binding domain connected to an effector domain with a linker (Bornholdt and Prasad, 2008). I attempted to determine the structure of ISAV-NS1 and compare it with the influenza NS1.

Similar to the ISAV-M1 and M2, the NS1 gene was inserted into a pET28b+ vector, and the ISAV-NS1 protein was successfully expressed in the Rosetta 2 cell line (Novagen). The ISAV-NS1 protein was purified with Ni-NTA resin, a HP SP column, and a size exclusion column. The final product of the NS1 protein was very pure and mono-dispersed. From the elution volume, the ISAV-NS1 was found to be a dimer in solution which was consistent with that of the influenza A virus NS1 protein. Crystal screens have been performed with around 300 commercial conditions, but no hit has been found. More ISAV-NS1 protein should be purified and a thorough screening of all available commercial conditions should be performed.
S1.4 ISAV-NEP

The ISAV-NEP protein is also encoded by gene segment 7, but the mRNA needs to be spliced before translation. The NEPs are important in the export of the viral RNPs, since they have nuclear export signals (NES) that enable the RNP-M1-NEP complexes to be recognized by the chromosome region maintenance 1 (CRM1) proteins which mediate the nuclear export from the nucleus. The crystal structure of C-terminal domain of the influenza NEP has been determined (Akarsu et al., 2003), which provides structural evidences on how NEP interacts with the influenza M1 protein. I attempted to determine the structure of the full length ISAV-NEP.

![Image of expression test for ISAV-NEP](image)

**Figure S1.6** Expression test for ISAV-NEP. W: whole cell lysate, S: supernatant fraction, P: cell pellet, Ft: flow through, E: eluted fractions. The red arrows point to the expected position for ISAV-NEP in SDS-PAGE, according to the molecular weight.

Similar to the other ISAV proteins, the NEP expression was tested in pET28b+ vector with an N-terminal His-tag fused. The NEP protein expression was induced with 0.1mM IPTG or 0.01mM IPTG at 15°C for 24 hours in the Rosetta 2 cell line (Novagen). However, the expression level of this protein and the solubility were not good when expressed in *E. coli*; the yield was low and most of the protein was in cell pellet (insoluble). Expression under different temperature and on-column refolding have also been tried, but the yield of soluble ISAV-NEP did not improve.
S2  Program for data analysis

To analyze the large amount of data obtained from RNA-protein binding measurement, I developed a program in Python to process data and perform statistical test. A graphical user interface (GUI) was later developed such that the other users do not have to know Python to use the program (Figure S2.1). This program was developed by using external packages such as numpy, scipy, and matplotlib. The GUI was constructed with the tkinter module. The function of this program was further extended to process large data set and perform model fitting and statistical test.

![Figure S2.1 The graphical user interface of the Data Analysis tool.](image)

The data can be read in from a “.csv” file (Figure S2.2). The program will do a data quality check to verify that the data is clean. Specifically, it checks whether there is any value that is scientifically impossible (e.g., a value falls out of the boundaries set by the user). The program also examines whether the standard deviation is bigger than the user-supplied value, and/or whether there are any extreme values (outliers). If the data is
good, the program will try applying different mathematical models and report the best one based on the fitting score what is related to root-mean-square deviation (RMSD). The modeled data and experimental data will be plotted together for visualization. Finally, statistical tests will be performed to show how well the model can explain the experimental data. Besides the data analysis functions, this tool can also be used to in simulation and forecasting, by supplying different experimental conditions.

**Figure S2.2** The data is read in and displayed in a data frame. Instructions are shown in the information windows.

**Figure S2.3** Modeling, plotting and statistical test. The output results are displayed in the info window.
S3  Genes and primer sequences

Proteins studied in this thesis project:

- ISAV NP
- ISAV NP ΔC16
- ISAV NP ΔN26
- ISAV NP ΔN39
- ISAV NP ΔN50
- ISAV NP ΔN64
- ISAV NP ΔN84
- ISAV NP ΔN111
- ISAV NP ΔN189
- ISAV NP ΔLOOP
- ISAV NP K185A/R186A
- ISAV NP K296A/R299A
- ISAV NP M1
- ISAV NP M2
- ISAV NP NS1
- ISAV NP NEP
**ISAV NP**

**Forward primer:** (NcoI)

TGAA CCATGG CA CATCATCATCATCATCATGCGATATAAAGGTATGA

**Reverse primer:** (XhoI)

CCGCTCGAGTTACTAAATGTCAATGTCTTCTCCTTC

---

178
ISAV NPΔC16

**Forward primer:** (NcoI)

```
TGAA CCATGG CA CATCATCATCATCATGCGATAAAAGTATGCAAGCAGACAACACTTGTTGTTGTAAG
```

**Reverse primer:** (XhoI)

```
CCGCTCGAGTTACTAAGGTTGTCATGTCAGTATTCT
```
ISAV NPΔN26

**Forward primer:** (NcoI) TGAA CCATGG CA CATCATCATCATCATGATCAAGTACCTCTCTCTCTCTCTCTCT

**Reverse primer:** (XhoI) CCGCTCGAGTTACTAATGTCATGTCATGTCATGTCATGTCATG
ISAV NpΔN39

ATG TAC CTT CTC CAA AAA AGG CAT TCG CCG GGA CAG AGA ATT GAA TTG GGT TGG GAG TTG AG

ATG TAC CTT CTC CAA AAA AGG CAT TCG CCG GGA CAG AGA ATT GAA TTG GGT TGG GAG TTG AG

Reverse primer: (XhoI)
CCG CTC GAG TTA CTA AAT GTC AAT GTC TTC CTC TTC CTC

Forward primer: (NdeI)
CCA ATT G CCA TAT G GGC AGC TCA CTT CTC CAA AAG GCA TTC GCC

CCA ATT G CCA TAT G GGC AGC TCA CTT CTC CAA AAG GCA TTC GCC
ISAV NPΔN50

**Forward primer:** (NdeI)

CCAATTG CATATG GGC AGC GAT GAA TTC TGG GTG GAG TAG

**Reverse primer:** (XhoI)

CCG CTCGAG TTA CTA AAT GTC AAT GTC TTC TGG GTG GAG TAG
**ISAV NPΔN64**

**ATG**
AAAAAGATTTAGAAAAATCTCTGGAAAGAGAGAAAAATGAAGGACATGAGACACAAGAGTGGTCTGGACAGCACTTGAAATGGAAGATAGATAGATGAGAAGGTGAAATGGAAGATAGATGAGAAGGGAGGTGCAGTGGTACCTGTGAAGGTGTCAGTGGTACCTGTGAAGGGAGGTGCAGTGGTACCTGTGAAG

**Forward primer:** (NdeI)
CCAATTG CATATG GGC AGC AAA AAG ATT AGA AAA TTC CTG GAA

**Reverse primer:** (XhoI)
CCG CTCGAG TTA CTA AAT GTC AAT GTC TTC TCT TTC TCT
**ISAV NPΔN84**

**Forward primer: (NdeI)**

```
ATGGGAGCGAGTGCGACGACGACGAATTTGGAAGATCGTTGAAATTTGACAAATTTTCACAAAAG
AAGCAGCGCTAATCAGTTGACTGGCAATTGAAATGCTGGTGTAGATGATGAAGAAGCTGGAGG
GTCTGGTAGACAAACAAAAGGGAAGACAAAGGGGCTCTCAAACATGGCCCTACAATCTG
CTATATTCATAGGGATGGTGTTTCCTGCTCTCACTACTTTCTTCAGTGCTATCCTATCAGA
AGGTGAAATGAGCATCTGCTGCAAAATGGACAAATCATGAGAATTCTGGCACTGGCA
GATGAAGACGGAAAGAGACAAACTCGGAGGAGAAGGTCCTCCTGACTAGCGCACTGAAAATCAG
CGAGGTTGAAAAACATGCCAGAGAAGCTGAGATGAAAGCATGAAAGGGGCAAGGTTCCAAG
GCTACAGAATCAAGCATCTCCAACCAATGGATATGGCACTGATTAT
GAAATCTGTGCTGTCAGCAGACCAACTTTTTGCTCCGGGAGTGAAGATGATGAGGACGGTT
TCAATGCGTCGTACACAACACTGGCAGAAGGGGCAAGGTTCCAAGGCTACAGAATCAAGCA
TCTCCAAGAATCTACTAAATGCTTCATCCACTATTTCCAGTCTCATGATGTACATCCA
AGAAGGCAAATCTGTGACTGGCAATGGGATATTTGCAATAAGACCGAGGAGAGCCCTGAAG
ATCTGCAAGAGGCAAGACCGGTAACACTGGCAACTTTTGCAAAATGCTAGGA
GACGGAACATCAGTGTTGCAACAGGCTTTTTGTTCTGCGAGATATCATC
GAGAATGCGAGTGCCTGTCCGAAAGGTCAGATCACAATCGCTAACAAACACAA
ACACAAAAAGGACCTTCCACATAACAAAAGGCTCAGTATAGTATGCTCAATCAGGAAGGTATG
AAGCAAGACCTGAAACACATAAAGACACAGACAGAAGTTGGCAAGGTTCAAGAGCTGAC
TGACAGCTGTGGAGACTTGCTGTTCAACAAGGGGCAATAGGAAGAAATGCTAGGA
GACGGAACATCAGAGAACAATCAGTGGAACGCTGACAGCCCTGGTGAAGAGAGAATTCTG
AATCACGTCCAGAGGAGTGCAAGAGCTGTAAAGACATTTATGGTGGGAACAAGG
GACAAATCAGTTTTGATCTGTGGTGGGAGATGTCTCATTGCAGCTTTTACAGAGGAGGTAG
AGGAAGCGGAAAGAAATACCTGACATGACACCTCGGAATTTGAGGTGAGCGGAGGACGCA
GGAAGAGGAAAGACATTTGACATTAG
```  

**Reverse primer: (XhoI)**

```
CTACTGCAAGAGGCAAGCTAATCAGTTGACTGGCAATTGAAATGCTGGTGTAGATGATGA
GAGAAGACGGAAAGAGACAAACTCGGAGGAGAAGGTCCTCCTGACTAGCGCACTGAAAATCAG
CGAGGTTGAAAAACATGCCAGAGAAGCTGAGATGAAAGCATGAAAGGGGCAAGGTTCCAAG
GCTACAGAATCAAGCATCTCCAACCAATGGATATGGCACTGATTAT
GAAATCTGTGCTGTCAGCAGACCAACTTTTTGCTCCGGGAGTGAAGATGATGAGGACGGTT
TCAATGCGTCGTACACAACACTGGCAGAAGGGGCAAGGTTCCAAGGCTACAGAATCAAGCA
TCTCCAAGAATCTACTAAATGCTTCATCCACTATTTCCAGTCTCATGATGTACATCCA
AGAAGGCAAATCTGTGACTGGCAATGGGATATTTGCAATAAGACCGAGGAGAGCCCTGAAG
ATCTGCAAGAGGCAAGACCGGTAACACTGGCAACTTTTGCAAAATGCTAGGA
GACGGAACATCAGTGTTGCAACAGGCTTTTTGTTCTGCGAGATATCATC
GAGAATGCGAGTGCCTGTCCGAAAGGTCAGATCACAATCGCTAACAAACACAA
ACACAAAAAGGACCTTCCACATAACAAAAGGCTCAGTATAGTATGCTCAATCAGGAAGGTATG
AAGCAAGACCTGAAACACATAAAGACACAGACAGAAGTTGGCAAGGTTCAAGAGCTGAC
TGACAGCTGTGGAGACTTGCTGTTCAACAAGGGGCAATAGGAAGAAATGCTAGGA
GACGGAACATCAGAGAACAATCAGTGGAACGCTGACAGCCCTGGTGAAGAGAGAATTCTG
AATCACGTCCAGAGGAGTGCAAGAGCTGTAAAGACATTTATGGTGGGAACAAGG
GACAAATCAGTTTTGATCTGTGGTGGGAGATGTCTCATTGCAGCTTTTACAGAGGAGGTAG
AGGAAGCGGAAAGAAATACCTGACATGACACCTCGGAATTTGAGGTGAGCGGAGGACGCA
GGAAGAGGAAAGACATTTGACATTAG
```
ISAV NPΔN111

**Forward primer: (NcoI)**

TGAA CCATGG CA CATCATCATCATCATCATGGTGTAGACATATGGA AGGGCTGGAGGAGTTGCTGAGACCTGGAACCAAGGAAAGAG

**Reverse primer: (XhoI)**

CCGCTCGAGTTAATAATTGCAATGCTGCTTCCTTC
ISAV NPΔN189

**ATG** ACAGGAGGACAGAGGTGGACATGGCTGATGTAACCAAGCTGAACGTAGTCACG
GCTAACGGGAAAGTCAAGCAAAGTGTGAACTAACCAGATCTCAAAAAGCAGCATTTCA
GGCAGAGTAGACCAAAGAGTCAAGAGCTACAGAAAGAAGGCAAGGTCTCCAGAGCTACAGA
ATCAAGCATCTCCAACCAATGTATTGTGACTTATGAGAATCTGCTGCTGCAAGCAGACC
AACATTGTGGTACCCAGAGGAAAGTCAAGATCTTTTCATCATTCAGAGGAAATCTGCGACACAGA
GGAGCAAGCGCTCTCTAGCTACGAGAAGTCGAGGGTAAAGAAAAACAGCAGGTGGGAACTGCCACAGA
ATGGGATGGAAGGCTTACGAGGTGGGAACTGCCACAGA
AACACTGGGAAAGGCTTACGAGGTGGGAACTGCCACAGA

**TAG**

Forward primer: (NcoI)
TGAA CCATGG CA CATCATCATCATCATCATCATACAGGAGGACAGAGGTGGAC

Reverse primer: (XhoI)
CCGCTCGAGTTTAAATGTTCAATGGCTTTCCTCTTC
ISAV NP ΔLOOP

**ATG**GGCGGATAAAAGGTGATCATTACTTCTTTTGATGTCAGAGACAACCTTGGTTGTAAG
AAGATCTACGGCTACTAAAAGTTGGCATTAAGATCCTCCTACAGAGAGGTGGAGCAACAT
CACCTTCAAAAAGGACTTCCGCAGGAGAGAAGATGAAATTTCTGGGTTGAGTTAGATCA
AGATGCTACGGCTACTAAAAGTTGGCATTAAGATCCTCCTACAGAGAGGTGGAGCAACAT
AAGATCTA
GGACATGGGCTGATGTAACCAAGCTGAACTGTCACGGCTAAACGGGAAAGTCAAGCAAA
GTTGAAGTAAACTGGAACCCCTCAACACTCCCTGACGGCAAGCAACATCAAACATG
GGACATGGGCTGATGTAACCAAGCTGAACTGTCACGGCTAAACGGGAAAGTCAAGCAAA
GTTGAAGTAAACTGGAACCCCTCAACACTCCCTGACGGCAAGCAACATCAAACATG

**Linker:** GGA GGA GGA GGA GGA TCC

**Forward primer:** (BamHI)
CGC GGATCC GGA GAC GGA ACA TCA GAG ACA ACT AGT

**Reverse primer:** (BamHI)
AAA GGA TCC TCC TCC TCC TCC TCC TCT GTC TGT GTT GCT TTG TGA TGT TTC
ISAV NPK185A/R186A

**Forward primer:**
AGAATTTCTGCCACTGGCAGATGAAAGCAGGAGGCAGCACAAACAAAGAAGACAGGAGGACAGA
GGGTGGAC

**Reverse primer:**
GTCCACCCCTCTGCTCCTCCTGTCTTGGTTTGTGGCTGCTCCGTCTCCATCTGACAGGAGCAGA
AATTCT
ISAV NPK296A/R299A

**Forward primer:**
ACACTGGCAGAAGGGCAGACATCCCGAGCGCTACCTAGCACACATGAGGAACTGCG
GAGGAGTAGCTCTGG

**Reverse primer:**
CAGAGCTACTCCTCGAGATTCAAGGAGTGCAAGAGCTGTAAAGACATTTATGGTGGAGAACAAGGGAAATCAGCTATTGTTATCTCTGGTGTGGGACTGTTCTCTATTGACTTTAAGGGGTAGAGGAAGCGGAAAGAATACTGACATGACACCTGAAATTGAGTTTGAGGACGAGGAGGAAGACATTGACATT
**ISAV-M1**

**ATG**

```
ACGAGATCATGAGAATACAAAAACATCTACCCTGAGATGAGAGAAGCAAACCCAAA
ACCACGAGGAGCGGATCAGAGGAAAGAAGAAGAAGGACCGGACTGAGGTGAGTTTAC
GCAATGGGAGTTCTCTCTGACACTGAGGGAGAGAGAGGACTGAGGCTGAAG
GTTCCAGAAGGAAAAACTTGAAAGGTCAAAAAACCTTGAAACCAATTGGTGAACGGAAATGA
TCAGTCGAGCGAGGTAGATCCCTACTCTGTGTGTGATGAAAGATACACATCCGAGGAAGCAT
GGCAACGCTGATGAGAACAGCACATCAGGAAACTCGTGAGAAGGAGGAATCAGGATGCAA
GGACCGGATGTGAGAGAGAAAAATGGCAATGTTGTGTTATGATTTTCACCCGA
CATGACGAGACAGAGCGATATGCTGAAGAGAGCTGAAAAACATGTCACACACGAGGAT
GCAGATGTATGCTCTGGGTCTGCAGAGTTCCATACAAAGGAGCTGAAAACAATTGCTGGAAC
GTCGATGCTCCTCCCTCCCTCGTCCAGAAGATGCTGAGTTCTCTGG
```

Forward primer: (NdeI)

```
CCAATTG CATATG ACGAGATCACAATGGATACAAAAACATCT
```

Reverse primer: (XhoI)

```
CCG CTCGAG TTA TCA CTTCCAGGTACCCAGAAGACCACATCTCTCTC
```
ISAV-M2

**ATG**

```
GATACAAAAACATCTACCATGACATGAGAGAAGCAAACCCAAAACCATGGGAGCTGA
TCAGACATGCCTTGAAGAGAAAAGAGAGGACGAGGTGGTTCGCAATGGGAGTTCT
TCTGACACTGGGGAGAGAGGACTGACCCTGGAGTTTCTGTTCCAGAAAGAAA
AACGTCGGAAGGTCAAAAACCTTTGAACCAATTTGATGAAACGAGGATCAGTCGAGCGACG
ATGACCCCTCTACTTGATGATGAAAGATCCACCACATCGGGAGAGCATGGAACGCTGATGA
GAGACCACATCAGGAACCTGGCTGAAGGAGGAATCAGGATGCCAGGACCGGATGCTG
GAGAGAAAAATGGGCAATGGGTGATTTAGGATTATGAAGAATGTCACTGCAGAGGAGGAGA
AGGAGATGCTGAAGGAGCTGAAGACATGCTACACAGCAGGATGCAATGTATGCTCT
GGGTGCAAGTTCGAAAGCCCTTAGAGAATTTAAGAAAAGCCATCGTCGCTGAGTATCG
CGACTTCCCAGCATCCTGCTGACAGAGAAGATGCTGCTTCTGGGTGTTACCTGAAATG
CGCTTCAAAGAAAGAATGGAAGCCGAGAAAGAAGAACTGAAAGAGCTTGGAGCCGCAAGAT
CTACAAGCTAAGAGAACATGGAAGGAGATGGAGTACAAGAAAATGAGGGATCAACCGA
GAAATCGACAAATTTGGAGGACTCTCTATACAA**TAA**
```

Forward primer: (NdeI)
CCAAATTG CATATG GATACAAAAACATCTACCATGACATGAGAGAAGCAAACCCAAAACCATGGGAGCTGA

Reverse primer: (XhoI)
CCG CTCGAG TTA TCA TTGTATAGAGTCTCCTCAATTGCTGATTTTC
ISAV-NS1

**ATG**GATTTCACCAAAAGTGATGGTGTGCTGGTTGACCAACTAAAACCTTCACGGAAAAAGA CAAAGTGCATAAAAATAGATTCTTGGGGGTCTGGCCTTTGAACATGTTGATTAATACAAAG AAAAAATGTCAGAACATGTCTGGGATTTAACCTGAGGTAATGGTGCCGGAACAAGGAGG AAAATGGTCTCTCCATGAAACGCGGGGTCATGTGTCCTGGTTTTACGGAGATGAT GAACCCAGGTGAAGGGTCCTGCGCAACTTCGTTGGAACATGTTGCATTTCCAAGTTGTTCC TCTGGGGAATGGGAGATGACTTCTTGCTCTGCTCGTGCGCTGGCCTAAGACACAAAAATGTTT GCCACACTTGCCCAGACGGATGACAAAGTTTCACTGACAGATTCATTGGAAGTTTTGTACTGACC AGAATGTCAGAACATGCTACTGACATTTGCCATGAAACTTTCGCTATTGCAGATG

Forward primer: (NdeI)
CCAATTGCATATG GATTTCACCAAAGTGTATGGTGCTGG

Reverse primer: (XhoI)
CCGCCTCGAGTTATCACTGACAGAAGCCTAGCGTCTACCATAGTGATGGGTGCACAGGG TCAGCAGCGAGGTTTCACTGACAGATTCATTGGAAGTTTTGTACCG
ISAV-NEP

ATGGATTTTCACCAAAAGTGATGTTGTTGCTGGTTGCCAACTAAAAACCTTCACGGGAAAGA
CAAGGTAGCTTCTTCTGTCGGACTCAAAAGGTTCTGGGGAGGATGGTACCTCAAGT
ACGTCAGGTATGCTGGGACCTTCTTGCAGGATCAAGTGGATTCATTTGCAATCAACGATTC
TACGACAGAGCCCCAAAACAAGACTGGATCCAGGGTTGTATCCATGGTTGAAATGGACG
GAGACCGCTTATCGCTCATCTACGAGAAGCCTAAGCTCTACATAGTGATGGGTGCACA
GGGTCAGCGAGCGAGGGTTCTGGAAAAAGGGATCAACATGAGAGACTGGAGTTGAGCATT
AGGGCTGGACCTCTCAGAAATGATGGATTGGTTGAAACTTGTTATGAAACAAGATTT
TTGTGTTTTTGTCAGAAAAAGAAAATTGCTGTAAACATGGAAGTTGAAAAATTCATTTGT
AATGAGAAACTTAA

Forward primer: (NdeI)
CCAATTG CATATG GATTTCACCAAAGTGATGTTGCTGGTT

Reverse primer: (XhoI)
CCG CTCGAG TTA TCA TTAGTTCATCATTACAAATGAATTTTTCAAC


of influenza A, B and C virus RNA segments are highly conserved and show partial inverted complementarity. Gene 8, 315-328.


199


the RNA-dependent polymerase encoding elements. EMBO J 8, 3867-3874.
maturation cofactor M2 is a cholesterol-binding protein. Eur Biophys J 34, 52-66.


forms oligomers and binds RNA. Nature 444, 1078-1082.