Structural Analysis of Influenza A Virus Nucleoprotein and its Interaction with RNA and Polymerase Subunit PB2

NG, Ka Leung

A Thesis Submitted in Partial Fulfilment

Of the Requirements for the Degree of

Doctor of Philosophy

in

Molecular Biotechnology

The Chinese University of Hong Kong

July 2011

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Thesis/ Assessment Committee

Professor TSUI Stephen Kwok Wing (Chair)
Professor SHAW Pang Chui (Thesis Supervisor)
Professor AU Shannon Wing Ngor (Committee Member)
Dr FODOR Ervin (External Examiner)

Acknowledgements

I would like to express my deepest thanks to my supervisor, Prof. Pang-Chui Shaw, for allowing me to work on this exciting project. I have been working under Prof. Shaw for almost seven years, since my undergraduate studies. I am grateful for his continuous guidance and support throughout these years, and especially a great deal of freedom in my research. I am very thankful for having a thoughtful and principled scientist to supervise my work.

Thanks are also devoted to Prof. Jia-huai Wang, who has supervised the crystallization work while I was in his laboratory in the United States for a total of ten months. I enjoyed discussing science with him and he has also shared with me his plentiful experience. Having a home-made dinner and watching the 2008 Beijing Olympic Opening Ceremony at his house was also memorable.

I would also like to thank the collaborators of this project: Hongmin Zhang and Jin-huan Liu for the structure of H5N1 NP; Zongli Li and Thomas Walz for the EM; Nicole Robb and Ervin Fodor for the RNP experiments; and Shannon Au for the molecular docking. I also thank my thesis committee, Prof. Shannon Au, Prof. Stephen Tsui and Dr. Ervin Fodor for their meaningful advices.

I am thankful to the Croucher Foundation for giving me the Research Studentship and Travel Grant. I also thank the CNOOC for granting me the Global Scholarship for Research Excellence for my visit to Prof. Jia-huai Wang's laboratory.

Thanks are given to my dearest labmates in the Shaw's group. Priscilla Too, Michelle Ho and Amanda Mak gave me guidance in my undergraduate research. Wai-Hon Chan, Mandy Lam, Yinhua Yang, Jessica Choi, Edwin Lo, May Li and Alice Poon worked with me in the influenza project. I would also like to thank Sue Law, Yuen-Ting Wong, Derek Cheung, Yiu-Ming Ng, Ming Li and Ka-lok Wong for their support.

I thank my parents and my younger brother, Andrew, for their love, care and support throughout these years. I also thank my friends for the happy and sad moments we shared. I am grateful to the brothers and sisters in Ka Fuk Baptist Church for their prayers and love.

Finally, I would like to give the honor and glory to the Lord Jesus Christ. I am thankful for His grace, presence and provision. It is great that He is the shepherd who I can put my faith and trust in. It is also my pleasure to know Him more through the science I am doing!

Abstract

The poultry-to-human transmission of the influenza virus and the recent H1N1 influenza pandemic have become major concerns worldwide. The nucleoprotein (NP) of influenza virus binds the RNA genome and plays essential role in transcription and replication during the virus life cycle.

We have determined the 3.3 Å crystal structure of H5N1 NP, which is composed of head and body domains and a tail loop. Using surface plasmon resonance (SPR), we found the basic loop (residues 73-91) and arginine-rich groove, but mostly a protruding element centering at R174 and R175, to be important in RNA binding. Ribonucleoprotein (RNP) reconstitution assay with these multiple-point and deletion mutants indicate their functional importance towards the transcription-replication activities of the virus polymerase. Single-point mutations at these concerned regions do not have a significant effect on their RNP activities, suggesting that NP mediates RNA-binding through multiple residues.

We have also shown, by RNP reconstitution assay and co-immunoprecipitation, that the interaction between NP and PB2 is crucial for the proper functioning of the RNP. The functional association of NP and PB2 requires either the PB2 host-determining residue lysine-627 or arginine-630 with the latter involving NP arginine-150 also. Using SPR, we have demonstrated that both residues take part in the direct protein-protein interaction, without the involvement of RNA. These results suggest a dual interaction mechanism between NP and PB2. This may confer

replication advantages to the virus, as either one can give an active RNP and explains the increased virulence of avian influenza viruses carrying the E627K mutation in mammalian cells. In addition, our findings identify the NP-PB2 interacting surface, with the PB2 627/630 region facing the RNA binding groove of NP.

The study leads to a better understanding towards the RNP organization of influenza virus and provides information for the future design of anti-influenza agents.

摘要

家禽傳人的流感病毒以及最近的 H1N1 流感大流行已成為全球重大的關注。流感病毒的核蛋白（NP）與其基因的 RNA 結合並在病毒生命週期的轉錄和複製過程中發揮重要作用。

我們已經確定了 H5N1 NP 的晶體結構至 3.3 Å，它由頭部和身體域和尾巴循環組成。利用表面等離子體共振（SPR），我們發現基本迴路（殘基 73-91）和精氨酸(Arginine)豐富的槽，但主要是一個以 R174 和 R175 為核心的突出元素，對 RNA 結合尤其重要。應用核糖核蛋白（RNP）的重組技術，這些多個檢測點和缺失突變體，表明了其功能對病毒聚合酶的轉錄-複製活動的重要性。單點突變在這些有關地區沒有顯著影響核蛋白的活動，這表明 NP 通過多個殘基介導其與 RNA 的結合。

通過重組核糖核蛋白法和免疫共沉澱，我們還表明了 NP 與 PB2 之間的相互作用對核蛋白的正常運作至關重要。NP 和 PB2 的功能結合，要麼需要 PB2 的宿主確定殘基賴氨酸(Lysine)- 627 或精氨酸- 630，而後者還涉及 NP 的精氨酸- 150。使用表面等離子體共振，我們已經表明，這兩個殘基直接參與蛋白質的相互作

用，RNA 並沒有牽涉其中。這些結果表明 NP 和 PB2 的雙相互作用機理。這可能賦予病毒複製的優勢，因為不論殘基 627 或 630 皆可形成一個正常運作的 RNP，這也解釋了 E627K 突變如何在哺乳動物細胞中增加禽流感病毒的毒力。此外，我們的調查結果也確定了 NP-PB2 相互作用的表面，是 PB2 殘基 627 及 630 地帶面對 NP 的 RNA 結合槽。

這項研究讓我們更好地理解流感病毒的核糖核蛋白的組織，並為今後設計抗流感藥物提供重要信息。

Table of Contents

Chapter 1
General Introduction

1.1 Severity of Influenza

Influenza is a contagious respiratory illness causing annual epidemics and occasional pandemics. The death toll of influenza epidemics is between 250,000 to 500,000 each year worldwide. Pandemics are highly deadly; for instance, the Hong Kong Flu in 1968 killed about 1 million people worldwide (Kilbourne, 2006). The most devastating Spanish Flu claimed the lives of 40 million people in 1918 (Palese, 2004). The relatively mild and recently ended H1N1 pandemic has resulted in over 18,449 deaths (World Health Organization, 2010). The 1997 influenza A H5N1 outbreak resulted in 6 deaths among 18 cases, with a high mortality rate of 33 % (Yuen et al., 1998). Starting from 2003, the poultry-to-human transmission of the virus has been reported from 15 countries in Asia and Africa. There are a total of 549 laboratory-confirmed human cases in which 320 are fatal (World Health Organization, 2011).

1.2 Introduction to Influenza Virus

1.2.1 Influenza Virus Particle and its Classification

Influenza virus belongs to the family of *Orthomyxoviridae*, which has a negative-sense, single-stranded and segmented RNA genome (Lamb and Krug, 2001). Dhori virus, Thogoto virus and Infectious salmon anemia virus are also within the same family. The virus particles are mainly spherical, with diameter ranging from

80 to 120 nm (Figure 1.1A). The budding virus particles are, however, elongated and sometimes filamentous (Figure 1.1B).

Influenza viruses are classified into types A, B and C according to the sequences of nucleoprotein (NP) and matrix protein (M1) (Lamb and Krug, 2001). Influenza virus types A and B possess eight segments of viral RNA (vRNA) while type C contains only seven (Desselberger et al., 1980). Influenza C viruses contain only one surface glycoprotein hemagglutinin-esterase-fusion (HEF) instead of the hemagglutinin (HA) and neuraminidase (NA) in Influenza A and B viruses.

Influenza A viruses are further classified into different subtypes based on the antigenic differences of the surface glycoproteins HA and NA, giving rise to H1N1, H5N1, H3N2, etc. There are currently 16 HA and 9 NA circulating in aquatic birds, only a few of them have been isolated from human (Fouchier et al., 2005; Laver et al., 1984). Unlike influenza A virus, there are no subtype divisions for influenza B virus, yet it is diverged into two antigenically distinct but cocirculating lineages, Victoria lineage and Yamagata lineage (McCullers et al., 2004; Rota et al., 1992).

1.2.2 Structure of Influenza A Virus

Influenza A viruses acquire a lipid envelope from the membrane of the host cells during budding off. From the electron micrograph, the glycoproteins HA and NA form rod-shaped and mushroom-shaped spikes on the surface of the envelope respectively and insert into the lipid bilayer. Small numbers of the M2 ion channels are embedded into the membrane. Underlying the lipid bilayer is the M1 protein, which is the most abundant protein in influenza virus and is thought to maintain the viral morphology (Bourmakina and Garcia-Sastre, 2003; Liu et al., 2002).

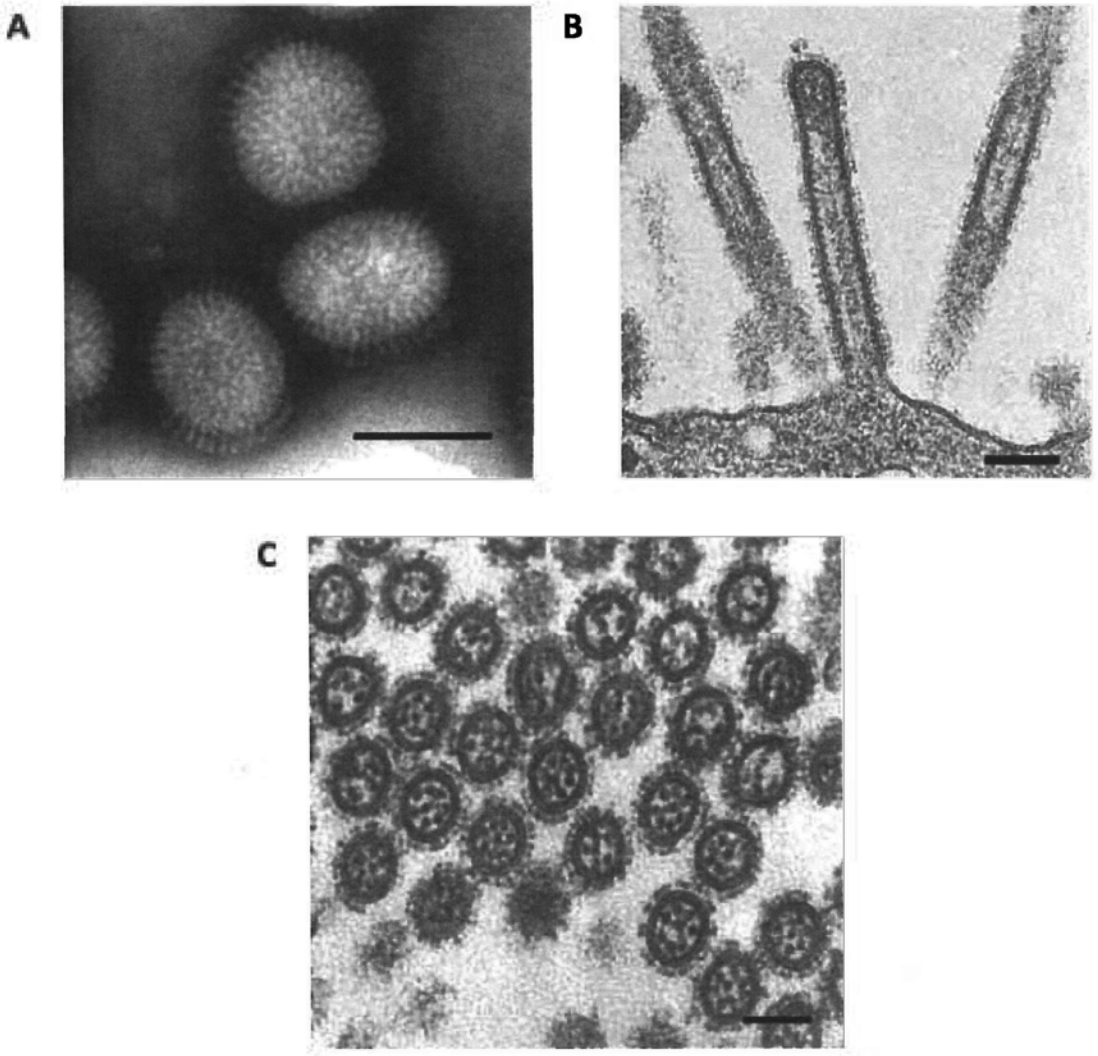


Figure 1.1 Electron microscopy of influenza A virus[1]. (A) Purified virions are in spherical shape of approximately 120 nm in diameter. (B) During budding, the virions are filamentous in shape. (C) Budding virions were observed by thin-section electron microscopy. The RNPs are organized in a '7+1' configuration. (Scale bar: 100 nm)

[1] Figure 1.1 is reprinted from Reviews in Medical Virology, Noda T. and Kawaoka, Y., Structure of influenza virus ribonucleoprotein complexes and their packaging into virions, 20:380-391, Copyright © (2010), with permission from John Wiley & Sons, Ltd.

Inside the virus, each of the eight segments of RNA is bound by NP, and a trimeric polymerase complex [acidic polymerase protein (PA), basic polymerase protein 1 (PB1) and basic polymerase protein 2 (PB2)], forming a viral ribonucleoprotein (vRNP) complex. The RNP complex was also found to associate with the M1 layer (Ye et al., 1999). Small amount of non-structural protein 2 (NS2), also known as nuclear export protein (NEP), is also found inside the virion (Richardson and Akkina, 1991; Yasuda et al., 1993). Two of the eleven influenza viral proteins, non-structural protein 1 (NS1) and pro-apoptotic protein PB1-F2, are not found in the virion but synthesized in the infected cells (Chen et al., 2001).

1.2.3 Infection Cycle

The infection cycle of influenza virus is summarized in Figure 1.2. During infection, influenza virus HA attaches to the sialylated receptor. Human viruses HA preferentially binds to the sialic acid receptors with an α2,6 linkage to galactose. Avian viruses HA, on the other hand, preferentially binds to those with an α2,3 linkage (Connor et al., 1994). The virus then enters the host cell through receptor-mediated endocytosis.

HA homotrimer in the virus particles contains two subunits, HA1 and HA2, which are linked through disulfide bond. It is noted that HA was synthesized as a precursor protein, HA0. Host cell proteases then cleaved HA0 into HA1 and HA2 in the secretory pathway (Klenk and Garten, 1994; Steinhauer, 1999). HA1 is for receptor binding as discussed above while HA2 is a transmembrane protein inserted into the viral membrane with an N-terminal fusion peptide.

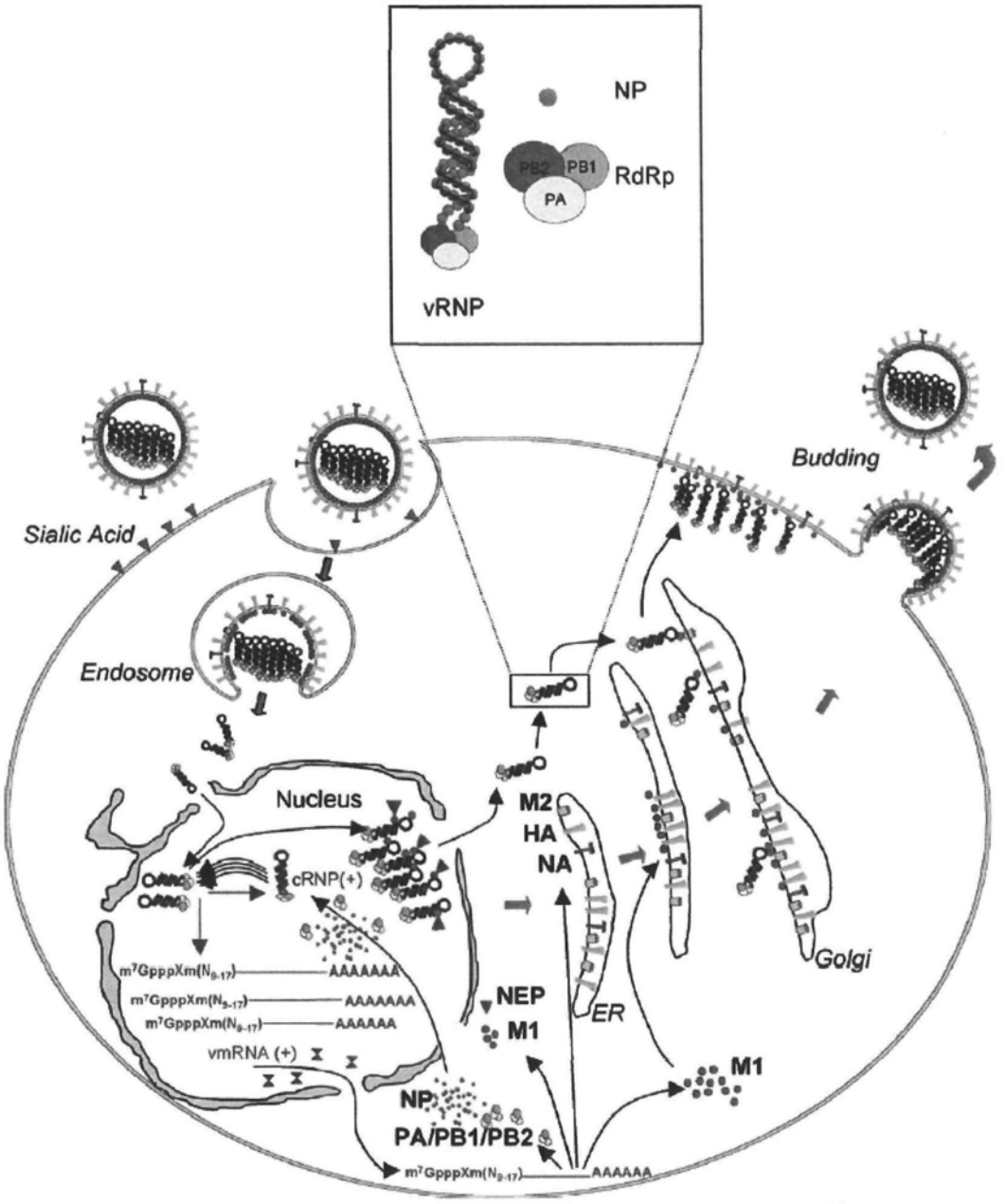


Figure 1.2 Infection cycle of influenza virus[2].

[2] Figure 1.2 is reprinted from Journal of Clinical Virology, 43, Josset, L., Frobert, E. and Rosa-Calatrava, M., Influenza A replication and host nuclear compartments: Many changes and many questions, 381-390, Copyright (2008), with permission from Elsevier.

The low pH of the endosome induces conformation change in HA2 and the N-terminal fusion peptide is inserted into the endosomal membrane, mediating the fusion process of the viral and endosomal membranes (Han et al., 2001; Skehel et al., 1982). The action of several HA trimers leads to fusion pore formation, and subsequently allows the RNP to exit the endosome and get into the cytoplasm. At the same time, the M2 ion channel permits the influx of hydrogen ions from the endosome to the virus particle. The low pH inside the virus dissociates vRNPs from the M1 layer and facilitates the exit of vRNP (Bui et al., 1996; Martin and Helenius, 1991). The eight vRNPs then enter the nucleus through the importin α-ß pathway. Evidences suggest that the nuclear localization signals (NLS) in NP play the crucial role in this process (Cros et al., 2005; O'Neill et al., 1995).

Inside the nucleus, negative-sense vRNAs are transcribed into positive-sense messenger RNAs (mRNA), which is 5' capped and 3' polyadenylated (Krug et al., 1979). These mRNAs are then exported into the cytoplasm and translated into different viral proteins using the host machinery. Those proteins that are essential for the formation of RNP, i.e. NP, PA, PB1 and PB2, are imported into the nucleus via their own NLS (Mukaigawa and Nayak, 1991; Nath and Nayak, 1990; Nieto et al., 1994). The vRNAs are also transcribed into positive-sense complementary RNAs (cRNA). cRNAs then associate with the newly imported NP and polymerase proteins and form the complementary RNP (cRNP), which serve as templates for synthesizing new vRNAs, and thus vRNPs, in the nucleus (Hay et al., 1982).

In the late phase of infection, M1 and NS2 are also imported into the nucleus. It is believed that vRNP is exported out of the nucleus through the Crm1-mediated pathway, which involves the formation of the vRNP-M1-NS2 complex, with NS2

further interacts with Crm1 (Neumann et al., 2000). HA, NA and M2 proteins get into the endoplasmic reticulum (ER) and the Golgi apparatus, where they are post-translationally modified (Doms et al., 1993). The three proteins are then sorted to the apical cell membrane, where virus assembly occurs (Kundu et al., 1996; Lin et al., 1998). M1 is thought to bring the vRNP and NS2 to the apical membrane for assembling (Bourmakina and Garcia-Sastre, 2005).

In the process of budding off, the eight vRNP are packaged into the virion. Two models have been proposed. The random incorporation model suggests that eight or more vRNPs are packaged randomly into the virion and not all virions are infectious (Bancroft and Parslow, 2002; Enami et al., 1991). The selective incorporation model suggests that each vRNA segment has a specific packaging signal and each virion possesses exactly eight vRNP (Duhaut and McCauley, 1996; Odagiri and Tashiro, 1997). Recent evidences have been more supportive of the selective incorporation model. The eight segment-specific packaging signals have been identified (reviewed in Noda and Kawaoka, 2010). Thin-section electron microscopy has revealed that the eight distinct vRNPs are organized into a 7+1 configuration (seven vRNPs surround a central core vRNP) in budding virions (Noda et al., 2006) (Figure 1.1C).

It is believed that M1 has the roles of fusing the membrane at the base of the virion and separating the virion from the cell membrane (Gomez-Puertas et al., 2000). Afterwards, the receptor-destroying activity of NA is required. NA cleaves the α-ketosidic linkage between the sialic acid and the carbohydrate on cell surface receptors and glycoproteins on the viral surface. This prevents HA from recognizing the sialic acid and therefore the virus particles would not aggregate with

each other or on the cell surface (Palese et al., 1974; Palese and Compans, 1976). The virion can then be successfully released from the infected cells.

1.3 Introduction to Ribonucleoprotein Complex

1.3.1 Structural Organization

vRNPs appear as supercoiled structures under electron microscopy (EM) (Compans et al., 1972, Pons et al., 1969). Due to the heterogeneity of the viral RNP, the EM image processing and 3D reconstruction have been largely dependent on a recombinant model-RNP. The model-RNP contains a circular ring of nine NP molecules bound with a model RNA and the trimeric polymerase PA, PB1 and PB2. The first reconstruction has reached the resolution of 27 Å at the NP ring and 36 Å at the polymerase (Martin-Benito et al., 2001) (Figure 1.3A).

The resolution of the polymerase was soon improved to 23 Å, with the domain positions of the polymerase subunits identified (Area et al., 2004). The C-terminal of PA and N-terminal of PB2 were revealed by the mass of their respective monoclonal antibody while the C-terminal of PB1 was indicated by the mass of the TAP-tag. Using cryo-EM, the resolutions at the NP ring and the polymerase were further boosted to 12 and 18 Å respectively, which allowed the fitting of the recently determined crystal structures of NP and the PA-PB1 complex into the 3D reconstruction (Coloma et al., 2009). This provides information for the polymerase subunit interaction and location of the vRNA.

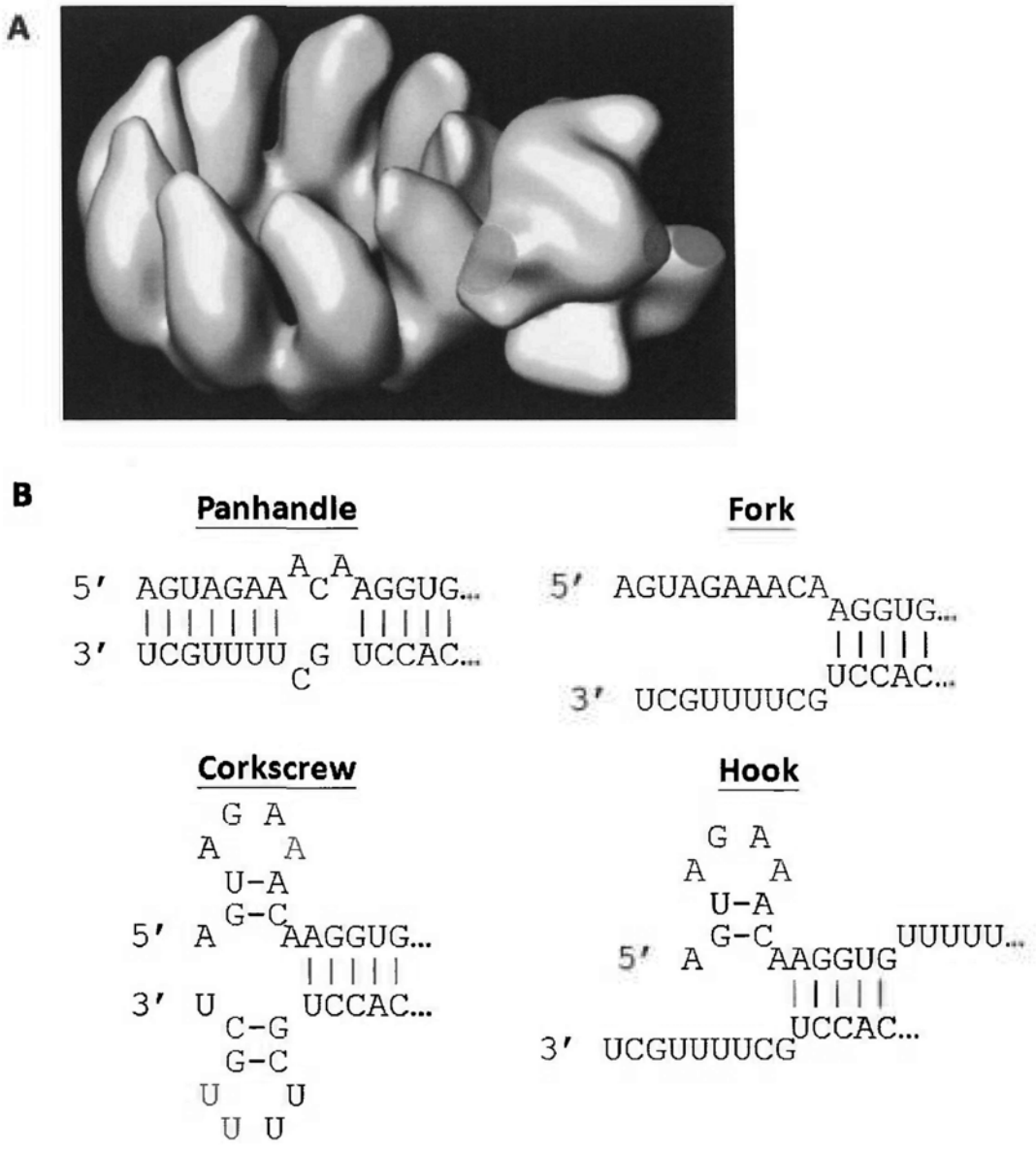


Figure 1.3 RNP and RNA structures[3]. **(A)** Three-dimensional reconstruction of RNP structure. The nine NP molecules are organized in a ring. Two of the NP molecules interact with the trimeric polymerase. Magneta region is the C-terminus of PA; green region is the C-terminus of PB1 while red region is the N-terminus of PB2. **(B)** Structural models of the vRNA promoter.

[3] Figure 1.3(A) is reprinted from Proceedings of the National Academy of Sciences of the USA, Area, E., Martin-Benito, J., Gastaminza, P., Torreira, E., Valpuesta, J. M., Carrascosa, J. L. and Ortin, J. 2004, 3D structure of the influenza virus polymerase complex: Localization of subunit domains, 101:308-313, Copyright © (2003) by The National Academy of Sciences of the USA

1.3.2 RNA

The eight vRNAs segments, numbered 1 to 8, are in descending order of lengths, which range from 890 to 2341 nucleotides. Part of the coding sequences contains the packaging signal for virion incorporation (Section 1.2.4). The 13 nucleotides at the 5' end and the 12 nucleotides at the 3' end are conserved among the 8 vRNAs and partially complementary to each other (Skehel and Hay, 1978; Robertson, 1979; Desselberger et al., 1980). Both the 5' and the 3' ends were shown to interact with the polymerase PB1 subunit (Gonzales and Ortin, 1999; Li et al., 1998). The pairing of the termini forms the panhandle structure, which is thought to result in the circular conformation of the vRNA (Hsu et al., 1987). It is also demonstrated that the double-stranded RNA element acts as the vRNA promoter. Besides the original panhandle structure formed by extensive base-pairing, several models were also proposed (Figure 1.3B).

The 'RNA-fork' model suggests that the extreme 5' and 3' ends do not form base-pairing but instead remain single-stranded. This is because defective mutations at these regions could not be restored by complementary double mutations (Fodor et al., 1994; Kim et al., 1997). The 'corkscrew' model suggests that the single-stranded regions in the 'RNA-fork' model form base-pairing within themselves (Flick et al., 1996). The 5' and 3' hairpin loops were found to stabilize and protect the polymerase against heat treatment (Brownlee and Sharps, 2002). The 3' hairpin loop was also found to be crucial for the endonuclease activity of the polymerase (Leahy et al., 2001). The 'hook' model suggests that only the extreme 5' terminus form a hairpin loop structure, which is required for the polyadenylation of the mRNA (Pritlove et al., 1999) (Section 1.3.4).

1.3.3 Trimeric Polymerase PA, PB1 and PB2

The trimeric polymerase is composed of three subunits: PA, PB1 and PB2. The C-terminal of PA was shown to bind the N-terminal of PB1, while the C-terminal of PB1 binds the N-terminal of PB2 (Gonzalez et al., 1996; Ohtsu et al., 2002) (Figure 1.4).

The function of PA has remained largely ambiguous in the past. It was shown to have proteolytic activity, but the activity does not correlate with the polymerase function (Naffakh et al., 2001; Sanz-Ezquerro et al., 1995). Its function has now been revealed after the resolution of the crystal structures. The endonuclease activity was previously thought to rely on PB1 subunit (Li et al., 2001), but the crystal structures clearly demonstrated that the PA N-terminal domain has similar folding and active site arrangement as the PD-(D/E)XK family of nuclease (Dias et al., 2009; Yuan et al., 2009). Mutagenesis experiments also confirmed the capability of the domain to hydrolyze single-stranded RNA and DNA (Dias et al., 2009). The C-terminal domain of PA instead is important for PB1-binding, as revealed from the co-crystal structures with the N-terminal peptide of PB1 (He et al., 2008; Obayashi et al., 2008). The PB1 peptide is clamped by the conserved cleft of PA, resembling the 'jaws' in the 'dragon head' (He et al., 2008).

PB1 is the core of the trimeric polymerase. Its main function is to catalyze the transcription of the three types of RNA (Braam et al., 1983). PB1 binds the 5' and 3' termini of vRNA (Section 1.3.2) and cRNA (Gonzalez and Ortin, 1999). The sequence of PB2 shows that it has conserved motifs of RNA-dependent RNA polymerases (Poch et al., 1989; Muller et al., 1994). The S-D-D motif (aa. 444-446) constitutes the active site for polymerization activity (Biswas and Nayak, 1994).

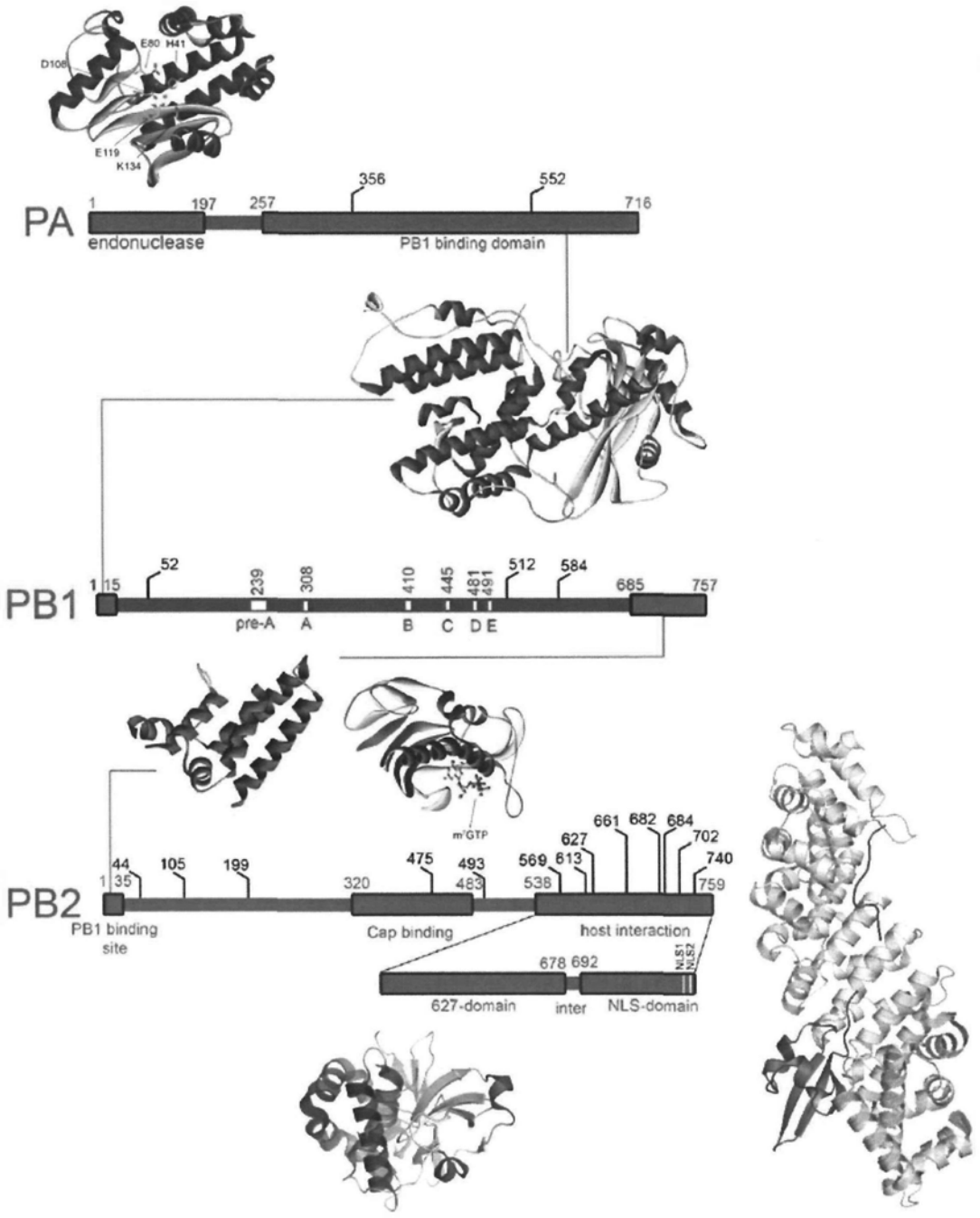


Figure 1.4 Structural organization of the trimeric polymerase[4]**.** All known structures are displayed according to the color code of the polymerase (Blue for PA, red for PB1 and green for PB2).

[4] This research was originally published in The Journal of Biological Chemistry. Boivin, S., Cusack, S., Ruigrok, R.W. and Hart, D.J. Influenza A virus polymerase: structural insights into replication and host adaptation mechanisms. *J Biol Chem*. 2010; 285: 28411-28417. © the American Society for Biochemistry and Molecular Biology.

The atomic structure of PB1 remains largely unknown, except for the N-terminal peptide (discussed above) and the C-terminal fragment (86 residues), which binds to the N-terminal of PB2 (37 residues). The PB1-PB2 co-crystal structure demonstrates a tight co-fold of these alpha helical regions (Sugiyama et al., 2009).

The major function of PB2 in the polymerase complex is to bind the cap structure of host pre-mRNA (Blaas et al., 1982; Ulmanen et al., 1981). Structural determination of PB2 domains has been hampered by the difficulty in expressing soluble constructs. It was until the development of ESPRIT (a robotic random gene fragment screening) that several PB2 domain structures were solved (Angelini et al., 2009). This includes the cap-binding domain (Guilligay et al., 2008), the '627-domain' (Kuzuhara et al., 2009; Tarendeau et al., 2008) (Section 1.5.4) and the NLS-domain (Tarendeau et al., 2007) (Section 1.5.2). The cap-binding domain (aa. 318-483) exhibits a novel fold; however, the binding mechanism of the methylated guanine base is similar to those in other cap-binding proteins, for example, eIF4E (Marcotrigiano et al., 1997). The aromatic ring of the base is sandwiched by the aromatic side chains of the protein.

1.3.4 Transcription of mRNA

mRNA transcription starts with the binding of 5' terminus of vRNA to PB1, which triggers PB2 to interact with the 5',7-methylguanosine cap structure of host pre-mRNAs (Cianci et al., 1995). The 3' terminus of vRNA is also bound by PB1, forming a duplex with the 5' end to stabilize the polymerase (Braam et al., 1983). The endonuclease domain of PA then cleaves the cap structure 9-15 nucleotides downstream, most likely after a purine residue (Beaton and Krug, 1981; Plotch et al., 1981). PB1 then uses this short piece of RNA as primer for mRNA transcription,

starting from the 3' of vRNA. It has been reported that a 'G' is added to the primer which corresponds to the second last residue 'C' of the 3' vRNA template (Beaton and Krug, 1981).

When mRNA synthesis approaches the 5' end of the vRNA template, a poly(A) tail has to be added. It is believed that the influenza polymerase copies the U-track of the vRNA at the 5' end repeatedly to generate a poly(A) tail for mRNA (Poon et al., 1999; Zheng et al., 1996). It is thought that the 5' hairpin loop assumes a 'hook' conformation (Figure 1.3B), which creates steric hindrance against PB1. PB1 then stutters on the preceding U-track and transcribes the poly(A) tail. The poly(A) tail is important for the nuclear export of mRNA (Poon et al., 2000).

1.3.5 Replication of vRNA

mRNA is an imperfect copy of vRNA, since the polymerase does not transcribe the 5' terminus, but instead adds the poly(A) tail. mRNA therefore lacks approximately 17 nucleotides and cannot act as a template for vRNA replication (Robertson et al., 1981). The polymerase transcribes cRNAs for that purpose. cRNA synthesis is a primer-independent process (Hay et al., 1982). cRNA is not prematurely terminated like the mRNA, and not polyadenylated. cRNA is encapsidated by NP after it is synthesized (Hay et al., 1977), and acts as the template for vRNA replication. The switch between mRNA transcription and vRNA replication has been poorly understood. Several models have pointed to NP for the regulation, which will be discussed in Section 1.3.6.

1.3.6 Nucleoprotein

The major role of NP in the RNP complex is to encapsidate vRNA and cRNA (Section

1.4.2). NP forms homo-oligomers to maintain the RNP structure (Prokudina-Kantorovich and Semenova, 1996) (Section 1.4.1), it is also thought to be the key adaptor for virus and host cell interaction (reviewed in Portela and Digard, 2002). NP interacts with PB1 and PB2 subunits in the viral RNA polymerase in forming the RNP (Biswas et al., 1998) (Section 1.4.3). It is thought to be the major switching factor that determines whether genomic vRNA is transcribed into mRNA or used as template to synthesize cRNA for genome replication (Skorko et al., 1991).

Several hypotheses have been put forward to explain the involvement of soluble NP in transcription regulation: (1) NP may act as a co-factor to co-transcriptionally coat the newly-synthesized cRNA (Shapiro and Krug, 1988); (2) NP may alter the structure of the RNA template to change its mode of synthesis (Fodor et al., 1994; Hsu et al., 1987; Klumpp et al., 1997); (3) NP may interact with polymerase subunits PB1 and PB2 to change the transcriptional activity of the polymerase (Biswas et al., 1998; Mena et al., 1999); (4) cRNA may be stabilized by the encapsidation of newly translated NP in the late phase of infection, while it may be degraded in the early phase (Vreede et al., 2004); (5) vRNA may be replicated by a soluble polymerase complex in trans while mRNA is transcribed by the cis-acting resident polymerase (Jorba et al., 2009); (6) small viral RNAs may interact with the polymerase and convert it from a transcriptase to a replicase (Perez et al., 2010). Besides, host factors interacting with the polymerase may also play some roles in regulating the replication process (Kawaguchi and Nagata, 2007).

1.4 Interaction of Nucleoprotein

1.4.1 NP-NP Homo-oligomerization

The homo-oligomerization of NP forms a major part of the RNP complex (Pons et al.,

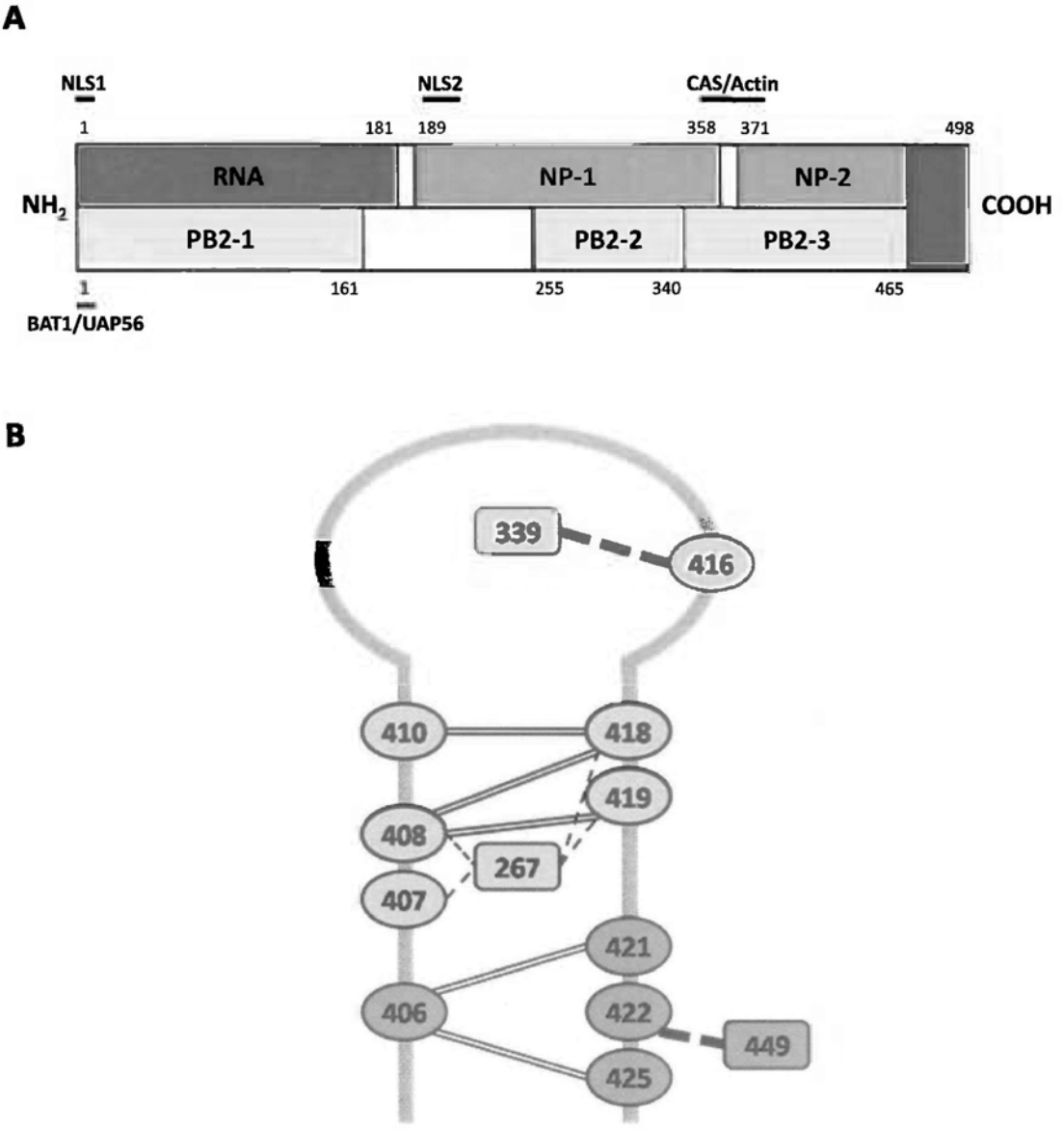


Figure 1.5 Schematic diagram of the interactions of NP. **(A)** Old view of the functional domains of NP before the structural determination. Our group has revisited the NP-NP interaction (green) and identified the crucial residues as shown in (B) (Chan et al., 2010). The NP-RNA (blue) and NP-PB2 (yellow) interactions have been revisited in this study. The red region was found to inhibit NP-NP and NP-PB2 interactions. **(B)** After the structural determination of NP, we have performed in vivo and in vitro studies to identify three forces for mediating NP-NP homo-oligomerization: (i) Interaction between the tail loop and insertion groove (blue); (ii) Maintenance of the tail loop conformation (red); and (iii) Stabilization of NP homo-oligomers (green).

1969), as previously observed for purified viral RNPs by EM (Ruigrok and Baudin, 1995) and for mini-RNPs (Area et al., 2004; Martin-Benito et al., 2001). Two positive elements at aa. 189-358 and 371-465 have been found to promote NP homo-oligomerization. A negative element at aa. 465-498 was also found to inhibit the interaction (Elton et al., 1999). However, NP homo-oligomerization and the transcription-replication activities were not well-correlated, since mutations in both positive (R199A and R416A) and negative (F479A) elements led to decreased transcriptional competence (Elton et al., 1999) (Figure 1.5A).

After the structural determination of NP, our group has re-investigated the essential residues for NP homo-oligomerization (Chan et al., 2010). Using an RNP reconstitution assay, we identified eight NP mutants that had different degrees of defects in forming functional RNPs, with the RNP activities of four mutants being totally abolished (E339A, V408S P410S, R416A, and L418S P419S mutants) and the RNP activities of the other four mutants being more than 50% decreased (R267A, I406S, R422A, and E449A mutants). Further characterization by static light scattering showed that the totally defective protein variants existed as monomers in vitro, deviating from the trimeric/oligomeric form of wild-type NP. The I406S, R422A, and E449A variants existed as a mixture of unstable oligomers, thus resulting in a reduction of RNP activity. Although the R267A variant existed as a monomer in vitro, it resumed an oligomeric form upon the addition of RNA and retained a certain degree of RNP activity. Our data suggest that there are three factors that govern the NP oligomerization event: (i) interaction between the tail loop and the insertion groove, (ii) maintenance of the tail loop conformation, and (iii) stabilization of the NP homo-oligomer (Figure 1.5B).

1.4.2 NP-RNA Binding

NP has been found to bind RNA with no sequence specificity but high affinity (Kd = 20 nM); NP also breaks the secondary structure of the RNA (Yamanaka et al., 1990; Baudin et al., 1994). Besides, chemical modification study showed that NP binds the phosphate backbone but not the nitrogenous bases of RNA (Baudin et al., 1994).

Studies have also been done to investigate where on NP binds RNA. The N-terminal 1-181 amino acids were found to be a RNA binding domain but the affinity was lower than the wild type protein (Kobayashi et al., 1994; Albo et al., 1995) (Figure 1.5A). UV crosslinking studies suggested that NP-RNA contact is mediated throughout the whole polypeptide (Elton et al., 1999a). They have identified arginine and tryptophan instead of lysine are responsible for RNA binding; and they discovered that several basic and aromatic amino acids including W120, W139, R267, W330, W386, F412 and R416 along the full length NP were important for RNA encapsidation (Elton et al., 1999a). It was also found that variants S314N and A332T were defective for RNA binding when expressed at the non-permissive temperature (Medcalf et al., 1999).

1.4.3 NP-Polymerase Binding

NP was found to interact with PB1 and PB2 but not PA, in recombinant system as well as virus-infected cells (Biswas et al., 1998; Medcalf et al., 1999). Three regions on NP (aa. 1-161, 255-340 and 340-465) were found to interact independently to PB2, while the C-terminal of NP (aa. 465-498) was found to inhibit NP-PB2 binding (Biswas et al., 1998) (Figure 1.5A).

1.4.4 Interaction between NP and Cellular Proteins

NP interacts with a number of cellular proteins, for the purpose of its localization in different compartments or its cellular functions (Figure 1.5A).

NP interacts with importin-alpha through its nuclear localization signals (NLS). Two NLSs have been well-characterized, the unconventional NLS-1 (aa. 1-13) (Wang et al., 1997) and the classical bipartite NLS-2 (aa. 198-216) (Weber et al., 1998). NLS-1 mainly mediates the nuclear import of NP and RNP while NLS-2 leads to nucleolar localization and is essential for virus replication (Ozawa et al., 2007). Recently, a third classical overlapping bipartite NLS was identified in some strains of NP (aa. 90-121) (Ketha and Atreya, 2008).

NP was shown to bind cytoskeleton in the late phase of infection (Avalos et al., 1997). It was later reported that NP bound F-actin in vitro and co-localized with ß-actin in vivo (Digard et al., 1999). A cytoplasmic accumulation signal (aa. 327-345) has been identified which targets NP to the actin cytoskeleton (Digard et al., 1999).

The shuttling of NP between the nucleus and the cytoplasm suggests that NP at some point is exported out of the nucleus (Neumann et al., 1997). NP was found to interact with CRM1/exportin-1 by in vitro assay. NP was also biased towards nuclear accumulation when the CRM1/exportin-1 pathway was blocked by leptomycin B (Elton et al., 2001).

Residues 1-20 of NP were also found to bind a cellular splicing factor, BAT1/UAP56, which belongs to the DEAD-box family of RNA-dependent ATPases. The interaction

was observed in vitro and in yeast-two-hybrid system. The interaction was thought to stimulate RNA synthesis of influenza virus (Momose et al., 2001).

Recently, NP was shown to directly interact with nuclear factor 90 by coimmunoprecipitation experiments. The two proteins colocalized in the nucleus in the early phase of infection. The interaction may exert a negative effect on the polymerase activity and virus replication of influenza virus (Wang et al., 2009).

1.5 Interaction of Polymerase Basic Protein 2

1.5.1 NP-PB2 Binding

Two PB2 fragments (N-terminal aa. 1-269 and C-terminal aa. 580-683) were identified to be responsible for NP-binding (Poole et al., 2004). There are, however, considerable overlaps of the NP- and PB1- binding sites on PB2. Two recent functional studies have suggested that NP-PB2 interaction is related to the host-determining residue K/E627 on PB2. They suggest that the reduced NP-polymerase and NP-PB2 binding of avian RNP in human cells and during the course of infection could be restored by the E627K mutation (Labadie et al., 2007, Rameix-Welti et al., 2009).

1.5.2 Interaction between PB2 and Cellular Proteins

PB2 has been shown to interact with several cellular proteins, for its folding and nucleocytoplasmic transport.

Heat shock proteins 70 and 90 were identified as interacting partners of PB2 (Hirayama et al., 2004; Momose et al., 2002). It has been proposed that the interaction may help the folding of PB2 (Chase et al., 2008), affect the nuclear

import of polymerase subunits (Naito et al., 2007) and influence the nuclear export of RNP (Hirayama et al., 2004).

Another better characterized partner is the cytosolic chaperonin containing TCP-1 (CCT) (Fislova et al., 2010). The central region of PB2 was found to interact with CCT. CCT was shown to play important roles in viral replication as well as vRNA and PB2 accumulation.

PB2 also interacts with importin α5 for mediating its nuclear import through a classical bipartite NLS (aa. 738-755). The crystal structure of the C-terminal domain of PB2 complexed with importin α5 has been determined (Tarendeau et al., 2007). The structure showed that the NLS unfolds upon binding to importin α5.

1.5.3 The Host-Determining Residue K/E627

The phenotypic differences between human and avian influenza viruses have been extensively studied throughout the decades. Human influenza viruses replicate more efficiently in mammalian cells than avian cells, and the same is true for avian viruses in avian cells compared to mammalian cells (Murphy et al., 1982). This host range restriction is conferred in part by PB2 (Naffakh et al., 2000, Mehle and Doudha, 2009), in which seventeen host-determining residues have been identified (Miotto et al., 2008). The best characterized residue, at position 627, is predominantly a lysine in human influenza viruses and a glutamate in avian influenza viruses. Avian polymerase with E627 was shown to be selectively restricted in human cells (Mehle and Doudha, 2008, Moncorge et al., 2010). Avian viruses with an E627K mutation have shown improved growth and enhanced virulence in infected mice and human cells. PB2 with K627 was also shown to

replicate more efficiently at 33°C (i.e. upper respiratory tract in human) than PB2 with E627 (Hatta et al., 2001; Mase et al., 2006; Massin et al., 2001).

1.5.4 Structure of the PB2 '627-domain'

The atomic structure of PB2 '627-domain' has been solved by two groups (Tarendeau et al., 2008; Kuzuhara et al., 2009). The N-terminal half of this domain (aa. 538-623) is alpha helical. This is followed by an extended peptide (aa. 624-634), which encircles helix alpha-5 of the N-terminal and contains the host-determining residue K/E627. The C-terminal half of the domain is composed of five beta-strands (aa. 635-675) and an extended loop (aa. 676-693) (Figure 1.4).

1.6 Objectives of the Study

The study is divided into three main parts. The first part is the structural determination of NP from avian origin. This was done through x-ray protein crystallography. The study provides molecular details on how NP performs its functions in the structural perspective. The structure aids the identification of possible inhibitor-binding sites and structure-based drug design.

The second and third parts are the investigation of NP-RNA and NP-PB2 interactions. These were done through a battery of biophysical, biochemical and cell-based experiments. The study identifies the determinants on NP which mediates RNA-binding activity and PB2-binding. The study also reveals how the two essential contacts are related to the transcription-replication process of influenza virus.

Chapter 2
Materials and Methods

2.1 Materials

2.1.1 Plasmids

Plasmid pRHisMBP was the gift from K.B. Wong (The Chinese University of Hong Kong). Plasmid pET28a was commercially purchased (Invitrogen, Carlsbad, CA). Plasmids pcDNA3a, pcDNA-PA (WSN and H5 origins), pcDNA-PB1 (WSN and H5 origins), pcDNA-PB2 (WSN and H5 origins), pcDNA-NP (WSN and H5 origins), pPOL1-NA-RT were the gifts from E. Fodor (University of Oxford) and have been previously described (Fodor et al., 2002; Vreede et al., 2004). pcDNA3.1/myc-His was the gift from K.F. Lau (The Chinese University of Hong Kong). pPOL-Luc-RT and pEGFP were the gifts from L.L.M. Poon (The University of Hong Kong) and have been described (Li et al., 2009).

2.1.2 Bacterial Cells

Competent cells DH5α, BL21(DE3)pLysS and BL21(DE3) were prepared according to the protocol in section 2.2.2.

2.1.3 Mammalian Cells

The 293T cell line (ATCC, Manassas, VA, USA) was cultivated in minimal essential medium (MEM) (Invitrogen, Carlsbad, CA, USA) with 10% fetal calf serum (Invitrogen).

2.1.4 Serum and Antibodies

Anti-NP serum was prepared by immunizing rabbits with purified NP. Anti-Myc antibody (Cell Signalling, Danvers, MA) and anti-beta-actin antibody (GenScript, Piscataway, NJ, USA) and anti-PB2 antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA) were purchased commercially.

2.1.5 Influenza Virus cDNA

cDNA of influenza virus A/HK/483/97(H5N1) was kindly provided by P.K.S. Chan (The Chinese University of Hong Kong).

2.2 Methods

2.2.1 Molecular Cloning

2.2.1.1 General Scheme

Influenza genes were amplified by specific primers using polymerase chain reaction (PCR) by Phusion Hot Start High Fidelity Polymerase (Finnzyme, Espoo, Finland) with 2720 Thermo Cycler (Applied Biosystems, Carlsbad, CA). Mutations were introduced by overlapping PCR. PCR products were purified by Gel-M Gel Extraction System (Viogene, Sijhih City, Taipei), according to manufacturer's instructions. Plasmids and purified PCR products were subject to restriction digestion by appropriate enzymes (New England Biolabs, Ipswich, MA, USA). Digested plasmid and PCR products were purified by Gel-M Gel Extraction System (Viogene), followed by ligation reaction using T4 DNA Ligase (New England Biolabs). Ligation products were transformed into *Escherichia coli* DH5α by calcium chloride method (Section 2.2.3). Transformed cells were spread on LB plate with appropriate antibiotics (100 μg/ml for ampicillin, 25 μg/ml for kanamycin, 50 μg/ml

for chloramphenicol) and incubated at 37 °C air bath overnight. Colonies were picked, inoculated into 5 mL LB medium with appropriate antibiotics, and incubated in 37 °C shaking air bath overnight at 250 rpm. Plasmids were extracted by Mini-Plus Plasmid DNA Extraction System (Viogene). Purified plasmids were subject to restriction digestion to confirm the presence of target gene. Cloned plasmids were sent for DNA sequencing (Tech Dragon Limited, Hong Kong, China) with either forward MBP sequencing primer (5' CGAGCTCGAACAACAACA 3'), forward standard T7 promoter primer or reverse standard Sp6 promoter primer.

2.2.1.2 Cloning of vectors for expressing NP in bacterial cells

Wild-type NP gene was cloned into pRHisMBP bacterial expression vector between the EcoRI and HindIII sites (pRHisMBP-NP-WT). The primers used for amplifying the full gene from the cDNA of influenza virus A/HK/483/97(H5N1) were H5N1NP-ECORI-F (5' CGCGAATTCGAATGGCGTCTCAAGGCACCAA 3') and H5N1NP-HINDIII-R (5' CGCAAGCTTTCAATTGTCATATTCCTCTG 3'). Mutations were introduced by overlapping PCR using the corresponding primers listed in Table 2.1 and the pRHisMBP-NP-WT as DNA template.

2.2.1.3 Cloning of vectors for expressing NP in mammalian cells

NP gene was cloned into pcDNA3a mammalian expression vector between the EcoRI and XhoI sites. The primers used for amplifying the full gene from the bacterial expression vectors (Table 2.1) were H5N1NP-ECORI-F (5' CGCGAATTCGAATGGCGTCTCAAGGCACCAA 3') and H5N1NP-XHOI-R (5' CGCCTCGAGTCAATTGTCATATTCCTCTG 3'). Clones and their templates are listed in Table 2.2.

Table 2.1 Primers for NP cloning (pRHisMBP vector)

Clone	Primer	Sequence (5' to 3')
pRHisMBP-NP-R74A	R74A-F	TCTGCATTTGATGAAGCAAGGAACAGGTACCTA
	R74A-R	TAGGTACCTGTTCCTTGCTTCATCAAATGCAGA
pRHisMBP-NP-R75A	R75A-F	GCATTTGATGAAAGGGCAAACAGGTACCTAGAG
	R75A-R	CTCTAGGTACCTGTTTGCCCTTTCATCAAATGC
pRHisMBP-NP-R74A,R75A	R74,75A-F	TCTGCATTTGATGAAGCAGCAAACAGGTACCTA
	R74,75A-R	TAGGTACCTGTTTGCTGCTTCATCAAATGCAGA
pRHisMBP-NP-R174A	R174A-F	GGATCAACCCTCCCGGCAAGATCTGGAGCTGCT
	R174A-R	AGCAGCTCCAGATCTTGCCGGGAGGGTTGATCC
pRHisMBP-NP-R175A	R175A-F	TCAACCCTCCCGAGGGCATCTGGAGCTGCTGGT
	R175A-R	ACCAGCAGCTCCAGATGCCCTCGGGAGGGTTGA
pRHisMBP-NP-R174A,R175A	R174,175A-F	GGATCAACCCTCCCGGCAGCATCTGGAGCTGCT
	R174,175A-R	AGCAGCTCCAGATGCTGCCGGGAGGGTTGATCC
pRHisMBP-NP-R221A	R221A-F	GCATATGAGGCAATGTGCAAC
	R221A-R	GTTGCACATTGCCTCATATGC
pRHisMBP-NP-G1 (R74A, R75A, R174A, R175A, R221A)	R74,75A-F	TCTGCATTTGATGAAGCAGCAAACAGGTACCTA
	R74,75A-R	TAGGTACCTGTTTGCTGCTTCATCAAATGCAGA
	R174,175A-F	GGATCAACCCTCCCGGCAGCATCTGGAGCTGCT
	R174,175A-R	AGCAGCTCCAGATGCTGCCGGGAGGGTTGATCC
	R221A-F	GCATATGAGGCAATGTGCAAC
	R221A-R	GTTGCACATTGCCTCATATGC
pRHisMBP-NP-R150A	R150A-F	GATGCCACATACCAGGCAACAAGAGCCCTCGTG
	R150A-R	CACGAGGGCTCTTGTTGCCTGGTATGTGGCATC

Table 2.1 (continued)

Clone	Primer	Sequence (5' to 3')
pRHisMBP-NP-R152A	R152A-F	ACATACCAGAGAACAGCAGCCCTCGTGCGGACT
	R152A-R	AGTCCGCACGAGGGCTGCTGTTCTCTGGTATGT
pRHisMBP-NP-R156A	R156A-F	ACAAGAGCCCTCGTGGCAACTGGAATGGACCCC
	R156A-R	GGGGTCCATTCCAGTTGCCACGAGGGCTCTTGT
pRHisMBP-NP-R162A	R162A-F	ACTGGAATGGACCCCGCAATGTGCTCTCTGATG
	R162A-R	CATCAGAGAGCACATTGCGGGGTCCATTCCAGT
pRHisMBP-NP-G2 (R150A, R152A, R156A, R162A)	R150,152A-F	GATGCCACATACCAGGCAACAGCAGCCCTCGTGCGGACT
	R150,152A-R	AGTCCGCACGAGGGCTGCTGTTGCCTGGTATGTGGCATC
	R156A-F	ACAAGAGCCCTCGTGGCAACTGGAATGGACCCC
	R156A-R	GGGGTCCATTCCAGTTGCCACGAGGGCTCTTGT
	R162A-F	ACTGGAATGGACCCCGCAATGTGCTCTCTGATG
	R162A-R	CATCAGAGAGCACATTGCGGGGTCCATTCCAGT
pRHisMBP-NP-G3 (K90A, K91A, K113A, R117A, R121A)	K90,91A-F	GCGGGGAAGGACCCGGCGGCGACCGGAGGTCCAATC
	K90,91A-R	GATTGGACCTCCGGTCGCCGCCGGGTCCTTCCCCGC
	K113,R117,R121A-F	CTGATTCTGTATGACGCAGAGGAGATAGCGAGAATTTGGGCCCAAGCGAACAATGGA
	K113,R117,R121A-R	TCCATTGTTCGCTTGGGCCCAAATTCTCGCTATCTCCTCTGCGTCATACAGAATCAG
pRHisMBP-NP-G4 (R382A, R384A)	R382,384A-F	AGCACTCTTGAACTGGCAAGCGCATATTGGGCTATAAGG
	R382,384A-R	CCTTATAGCCCAATATGCGCTTGCCAGTTCAAGAGTGCT
pRHisMBP-NP-Δ74-88	DEL74-88-F	CCGAAGAAGACCGGAGG
	DEL74-88-R	CCTCCGGTCTTCTTCGGTTCATCAAATGCAGAG

2.2.1.4 Cloning of vectors for expressing myc-tagged NP in mammalian cells

NP gene was cloned into pcDNA3.1/myc-His vector between EcoRI and XhoI sites for the expression of myc-tagged NP in mammalian cells. The primers used for amplifying the full gene from the bacterial expression vectors (Table 2.1) were H5N1NP-ECORI-F (5' CGCGAATTCGAATGGCGTCTCAAGGCACCAA 3') and H5N1NP-XHOI-R(MYC) (5' CGCCTCGAGATTGTCATATTCCTCTG 3'). Clones and their templates are listed in Table 2.3.

2.2.1.5 Cloning of vectors for expressing PB2 in mammalian cells

Domain-swapped PB2 mutants were cloned into pcDNA3a mammalian expression vector between BamHI and XhoI sites. Individual PB2 DNA fragments from WSN(H1) or H5 origins were first amplified and were used as template for second-round PCR. Primers used in the second-round PCR were either H5PB2_1_BAMHI_F (5' CGCCGCGGATCCATGGAGAGAATAAAAGAACTAAG 3') or H1PB2_1_BAMHI_F (5' CGCCGCGGATCCATGGAAAGAATAAAAGAACTAAG 3'); and PB2_759_XHOI_R (5' ATCTCACTCGAGCTAATTGATGGCCATCCGAAT 3') (the reverse primer was shared by WSN(H1) and H5 PB2). First-round primer sequences and templates specific for the clones are listed in Table 2.4.

H5 PB2 point mutants were cloned into pcDNA3a mammalian expression vector between the BamHI and XhoI sites. The primers used for amplifying the full gene from the pcDNA-PB2(H5) were H5PB2_1_BAMHI_F (5' CGCCGCGGATCCATGGAGAGAATAAAAGAACTAAG 3') and PB2_759_XHOI_R (5' ATCTCACTCGAGCTAATTGATGGCCATCCGAAT 3'). Mutations were introduced by overlapping PCR using the corresponding primers listed in Table 2.5.

Table 2.2 Templates for NP cloning (pcDNA3a vector)

Clone	Template
pcDNA-NP-WT	pRHisMBP-NP-WT
pcDNA-NP-R74A	pRHisMBP-NP-R74A
pcDNA-NP-R75A	pRHisMBP-NP-R75A
pcDNA-NP- R74A,R75A	pRHisMBP-NP- R74A,R75A
pcDNA-NP-R174A	pRHisMBP-NP-R174A
pcDNA-NP-R175A	pRHisMBP-NP-R175A
pcDNA-NP- R174A,R175A	pRHisMBP-NP- R174A,R175A
pcDNA-NP-R221A	pRHisMBP-NP-R221A
pcDNA-NP-G1	pRHisMBP-NP-G1
pcDNA-NP-R150A	pRHisMBP-NP-R150A
pcDNA-NP-R152A	pRHisMBP-NP-R152A
pcDNA-NP-R156A	pRHisMBP-NP-R156A
pcDNA-NP-R162A	pRHisMBP-NP-R162A
pcDNA-NP-G2	pRHisMBP-NP-G2
pcDNA-NP-Δ74-88	pRHisMBP-NP-Δ74-88

Table 2.3 Templates for myc-tagged NP cloning (pcDNA3.1/myc-His vector)

Clone	Template
pcDNA3.1/myc-His-NP-WT	pRHisMBP-NP-WT
pcDNA3.1/myc-His-NP-G1	pRHisMBP-NP-G1
pcDNA3.1/myc-His-NP-G2	pRHisMBP-NP-G2
pcDNA3.1/myc-His-NP-Δ74-88	pRHisMBP-NP-Δ74-88
pcDNA3.1/myc-His-NP-R150A	pRHisMBP-NP-R150A

Table 2.4 Primers and templates for domain-swapped PB2 cloning (pcDNA3a vector)

Clone	Primer	Sequence (5' to 3')	Template
pcDNA-PB2-H5[1-158]H1[159-759]	PB2_159_F	TTCCCAAATGAAGTTGGAGCCAGG	pcDNA-PB2(WSN)
	PB2_158_R	CCTGGCTCCAACTTCATTTGGGAA	pcDNA-PB2(H5)
pcDNA-PB2-H5[1-279]H1[280-759]	PB2_280_F	GCACACAAATTGGCGGAGTAAGAATGG	pcDNA-PB2(WSN)
	PB2_279_R	CCATTCTTACTCCGCCAATTTGTGTGC	pcDNA-PB2(H5)
pcDNA-PB2-H5[1-438]H1[439-759]	PB2_439_F	AATCAACGACTGAACCCCATGCACC	pcDNA-PB2(WSN)
	PB2_438_R	GGTGCATGGGGTTCAGTCGTTGATT	pcDNA-PB2(H5)
pcDNA-PB2-H5[1-550]H1[551-759]	PB2_551_F	ATAACTTATTCATCGTCTATGATGTG	pcDNA-PB2(WSN)
	PB2_550_R	CACATCATAGACGATGAATAAGTTAT	pcDNA-PB2(H5)
pcDNA-PB2-H1[1-279]H5[280-759]	PB2_280_F	GCACACAAATTGGCGGAGTAAGAATGG	pcDNA-PB2(H5)
	PB2_279_R	CCATTCTTACTCCGCCAATTTGTGTGC	pcDNA-PB2(WSN)
pcDNA-PB2-H1[1-158]H5[159-759]	PB2_159_F	TTCCCAAATGAAGTTGGAGCCAGG	pcDNA-PB2(H5)
	PB2_158_R	CCTGGCTCCAACTTCATTTGGGAA	pcDNA-PB2(WSN)

H1 PB2 point mutants were cloned into pcDNA3a mammalian expression vector between the BamHI and XhoI sites. The primers used for amplifying the full gene from the pcDNA-PB2(WSN) were H1PB2_1_BAMHI_F (5' CGCCGCGGATCCATGGAAAGAATAAAAGAACTAAG 3') and PB2_759_XHOI_R (5' ATCTCACTCGAGCTAATTGATGGCCATCCGAAT 3'). The K627E mutations were introduced by overlapping PCR using H1_K627E-F (5' CGCTCCACCAGAACAAAGTGG 3') and H1_K627E-R (5' CCACTTTGTTCTGGTGGAGCG 3').

2.2.1.6 Cloning of vectors for expressing PB2 '627-domain' in bacterial cells

H5 PB2 '627-domain' (aa. 538-693) was cloned into pET28a bacterial expression vector between the NheI and HindIII sites. The primers used for amplifying the full gene from the mammalian expression vectors (Table 2.5) were BAC_H5PB2_538-693_NHEI_F (5' GTTCTCGCTAGCGAAATTAACGGCCCAGAATC 3') and BAC_H5PB2_538-693_HINDIII_R (5' GACTATAAGCTTTTACCCTCTCAATACTGCAGAC 3'). Clones and their templates are listed in Table 2.6.

2.2.1.7 Cloning of vectors for expressing myc-tagged polymerase subunits in mammalian cells

H5 PA, PB2 and WSN(H1) PB2 were cloned into pcDNA3.1/myc-His vector between KpnI and XhoI sites for the expression of myc-tagged polymerase subunits in mammalian cells. The primers and templates used for amplifying the full gene are listed in Table 2.7.

2.2.2 Preparation of Competent Cells

Bacterial cells were streaked onto LB (DH5α and BL21(DE3)) or LBC (BL21(DE3)pLysS)

Table 2.5 Primers for point mutant PB2 cloning (pcDNA3a vector)

Clone	Primer	Sequence (5' to 3')
pcDNA-PB2-E567N	PB2_567_F	GGTCCCAAAATCCCACCATGC
	PB2_567_R	GCATGGTGGGATTTTGGGACC
pcDNA-PB2-A588V,S590G	PB2_588,590F	CCAAAGGCTGTTAGAGGCCAATATAGTG
	PB2_588,590R	CACTATATTGGCCTCTAACAGCCTTTGG
pcDNA-PB2-V613A	PB2_613F	CGTTTGACACTGCTCAAATAATC
	PB2_613R	GATTATTTGAGCAGTGTCAAACG
pcDNA-PB2-E627K	H5_E627K-F	GCCCCACCTAAGCAGAGTAGG
	H5_E627K-R	CCTACTCTGCTTAGGTGGGGC
pcDNA-PB2-R630G	H5_R630G-F	GAACAGAGTGGGATGCAATTTTC
	H5_R630G-R	GAAAATTGCATCCCACTCTGTTC
pcDNA-PB2-E627K,R630G	PB2_627,630F	CCCCACCTAAGCAGAGTGGAATGCAATTT
	PB2_627,630R	AAATTGCATTCCACTCTGCTTAGGTGGGG
pcDNA-PB2-V638I	PB2_638_F	CTCTGACTATAAATGTGAGGGG
	PB2_638_R	CCCCTCACATTTATAGTCAGAG
pcDNA-PB2-A655V	PB2_655F	GTAACTCCCCTGTATTTAACTAC
	PB2_655R	GTAGTTAAATACAGGGGAGTTAC
pcDNA-PB2-I667V	PB2_667_F	GAGGCTTACAGTTCTTGGGAAGG
	PB2_667_R	CCTTCCCAAGAACTGTAAGCCTC
pcDNA-PB2-A674P	PB2_674F	GGACGCAGGTCCTCTTACAGAG
	PB2_674R	CTCTGTAAGAGGACCTGCGTCC
pcDNA-PB2-K702R	PB2_702_F	GCAAAGAAGACAGGAGATATGGAC
	PB2_702_R	GTCCATATCTCCTGTCTTCTTTGC
pcDNA-PB2-T717A	PB2_717_F	GAGCAATCTTGCGAAAGGGGAG
	PB2_717_R	CTCCCCTTTCGCAAGATTGCTC
pcDNA-PB2-D740N	PB2_740F	CGGAAACGGAACTCTAGCATA
	PB2_740R	TATGCTAGAGTTCCGTTTCCG

Table 2.6 Templates for PB2 '627-domain' cloning (pET28a vector)

Clone	Template
pET28a-PB2-WT	pcDNA-PB2-WT
pET28a-PB2-E627K,R630G	pcDNA-PB2-A588V,S590G
pET28a-PB2-R630G	pcDNA-PB2-R630G

Table 2.7 Primers and Templates for myc-tagged polymerase subunits cloning (pcDNA3.1/myc-His vector)

Clone	Primer	Sequence (5' to 3')	Template
pcDNA3.1/myc-His-PA	H5_PA_myc_KpnI_F	ATTGCTGGTACCATGGAAGACTTTGTGCGACAA	pcDNA-PA(H5)
	H5_PA_myc_XhoI_R	AGCAATCTCGAGTCTTAGTGCATGTGTGAGGAA	
pcDNA3.1/myc-His-PB2(H5)	H5_PB2KpnIF	CGCGCCGGTACCATGGAGAGAATAAAAGAACTAAG	pcDNA-PB2(H5)
	H5H1_PB2XhoIR	CGCAGACTCGAGATTGATGGCCATCCGAATTCT (shared by WSN and H5 PB2)	
pcDNA3.1/myc-His-PB2(H1)	H1_PB2KpnIF	CGCGCCGGTACCATGGAAAGAATAAAAGAACTAAGG	pcDNA-PB2(WSN)
	H5H1_PB2XhoIR	CGCAGACTCGAGATTGATGGCCATCCGAATTCT	

plates and incubated at 37 °C overnight to obtain a single colony. The colony was picked, inoculated into 5 mL LB or LBC medium and incubated at 37 °C overnight with constant shaking at 250 rpm. 0.5 mL of the bacterial culture was added to 100 mL LB or LBC medium and incubated at 37 °C with constant shaking at 250 rpm until OD_{600} reached 0.45. The cells were then kept on ice for 5 minutes, and centrifuged at 4,000 g for 10 minutes at 4 °C. The cell pellet was resuspended in 40 mL RF1 buffer (50 mM KAc, 100 mM $RbCl_2$, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15 % glycerol, pH 5.8 [adjusted by acetic acid], filter-sterilized and wrapped with aluminium foil). The suspension was kept on ice for 5 minutes and centrifuged again at 4,000 g for 10 minutes at 4 °C. The cell pellet was resuspended in 4 mL RF2 buffer (10 mM 3-(N-morpholino) propanesulfonic acid (MOPS), 75 mM $CaCl_2$, 10 mM $RbCl_2$, 15 % glycerol, pH 6.5 [adjusted with KOH], autoclaved). The cell suspension was kept on ice for 15 minutes, and made into aliquots of 100 µl/tube. The competent cells were frozen by liquid nitrogen and stored at -80 °C.

2.2.3 Transformation by Heat Shock

0.5 µl of plasmid DNA was added into 100 µl of freshly thawed competent cells. It was then kept on ice for 30 minutes. Heat shock was performed at 42 °C for 2 minutes. The tube was returned to ice for 10 minutes. 500 µl LB medium was added into the tube and incubate at 37 °C air bath with constant shaking at 250 rpm for 45 minutes for recovery.

2.2.4 Expression and Purification of NP

Maltose binding protein (MBP)-tagged NP was expressed in *Escherichia coli* BL21(DE3)pLysS. Transformed cells were grown in LB medium with 100 µg/ml ampicillin and 50 µg/ml chloramphenicol until OD_{600} reached 0.6-0.8. Protein

expression was then induced by 0.4 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) at 25 °C for 16 hours with constant shaking at 250 rpm. Cells were harvested by centrifugation at 8,000 g for 5 minutes at 4 °C.

The cells were lysed by sonication, and the lysate was centrifuged at 16,000 g for 1 hour at 4 °C. The supernatant was passed through an amylose column (New England Biolabs). Bound protein was eluted with a 0-20 mM maltose gradient in 20 mM sodium phosphate (pH 6.5) and 150 mM NaCl. The eluate was incubated with thrombin (100 U) (Sigma, St. Louis, MI, USA) and RNaseA (100U) (Sigma) at 4°C overnight to remove MBP from NP and then passed through a heparin HP column (GE Healthcare). NP was eluted with a 0-1.5 M NaCl gradient in the same buffer. Gel filtration was performed with Superdex 200 (GE Healthcare) in 20 mM MOPS, 150 mM NaCl, pH 7.0. RNaseA was removed after passing through heparin HP column and gel filtration. Purified NP was concentrated to about 10 mg/ml with Amicon Ultra-15 Centrifuge Filter Units with 10,000 MWCO (Millipore, Billerica, MA, USA). Concentrated proteins were frozen by liquid nitrogen and stored at -80 °C.

2.2.5 Expression and Purification of PB2 '627-domain'

The method has been described previously (Tarendeau et al., 2008). His-tagged PB2 '627-domain' was expressed in *Escherichia coli* BL21(DE3). Transformed cells were grown in LB medium with 25 μg/ml kanamycin until OD_{600} reached 0.8-1.0. Protein expression was induced by 0.4 mM IPTG at 25 °C for 5 hours with constant shaking at 250 rpm. Cells were harvested by centrifugation at 8,000 g for 5 minutes at 4 °C.

The cells were lysed by sonication, in lysis buffer (30 mM Tris-HCl, 200 mM NaCl, pH 7.0). The lysate was centrifuged at 16,000 g for 1 hour at 4 °C. The supernatant was then passed through a His column (GE Healthcare). Bound protein was washed with lysis buffer, 1 M NaCl, 50 mM imidazole and 75 mM imidazole sequentially to reduce non-specific binding. His-tagged PB2 '627-domain' was eluted with 500 mM imidazole in lysis buffer. Gel filtration was then performed with Superdex 75 (GE Healthcare) in lysis buffer or in phosphate buffered saline (PBS). Purified PB2 '627-domain' was concentrated to about 10 mg/ml with Amicon Ultra-15 Centrifuge Filter Units with 10,000 MWCO (Millipore). Concentrated proteins were stored at 4 °C.

2.2.6 Static Light Scattering

RNase-treated NP and untreated NP, both at a concentration of 2.5 mg/ml, were subject to static light scattering analysis using a miniDAWN triangle (45, 90, 135°) light-scattering detector (Wyatt Technology Corporation, Santa Barbara, CA, USA) connected to an Optilab DSP interferometric refractometer (Wyatt Technology Corporation). This system was connected to a Superdex 10/300 GL Column (GE Healthcare), controlled by an ÅKTAexplorer chromatography system (GE Healthcare). Before sample injection, the miniDAWN detector system was equilibrated with 20 mM sodium phosphate and 150 mM NaCl, pH 7.0, for at least 2 hours to ensure a stable baseline signal. The flow rate was set to 0.7 ml/min, and the sample volume was 100 µl. The laser scattering (687 nm) and the refractive index (690 nm) of the respective protein solutions were recorded during the measurement processes. Wyatt software ASTRA was used to evaluate all data obtained.

2.2.7 Electron Microscopy

A 51-nt RNA (5' GGG AGA UUU UUU UUU UUU UUU UUU UUU UUU UUU UUU UUU UUU UUU UUU UUU UUU UUU 3') was produced by in vitro transcription using the MEGAscript kit (Ambion, Austin, TX, USA), purified by phenol extraction, and mixed with NP at a molar ratio of 1:2 in buffer containing 0.1 M sodium phosphate and 0.15 M NaCl, pH 6.5. EM samples of NP with or without RNA were prepared by conventional negative staining as described (Ohi et al., 2004). Briefly, 3.5-µl sample was applied to a glow-discharged carbon-coated EM grid. After 30 second incubation, the grid was washed with 2 drops of Milli-Q water (Millipore) and stained with 2 drops of 0.75% (w/v) uranyl formate. Grids were examined with a Philips CM10 electron microscope (Philips Electronics, Mahwah, NJ, USA) equipped with a tungsten filament and operated at an acceleration voltage of 100 kV. Images were recorded with a Gatan 1 x 1 k CCD camera (Gatan, Inc., Pleasanton, CA, USA) at x52,000 and a defocus value of ~2 µm.

2.2.8 Crystallization of NP

RNase-treated NP in 20 mM MOPS, 100 mM NaCl, pH 7.0, was first crystallized by vapor diffusion using the sitting drop technique. Initial crystallization conditions were surveyed using Index, Crystal screen I/II, and Natrix kits (Hampton Research, Aliso Viejo, CA, USA). Small crystals were obtained in Natrix condition no. 5 [0.2 M KCl, 0.01 M $MgCl_2$, 0.05 M 2-(N-morpholino)ethane sulfonic acid, pH 5.6; and 5% PEG 8000]. After refinement of the crystallization conditions, tetrahedral crystals were grown in hanging drops in 0.15 M KCl, 0.01 M $MgCl_2$, 0.1 M sodium cacodylate, pH 6.2, and 7% PEG 8000.

2.2.9 Structural Determination of NP

Crystals were soaked in increasing concentrations of sucrose up to 20% in the crystallization buffer. Frozen crystals were brought to Argonne National Laboratory for data collection. The data were processed and scaled with the HKL3000 suite (Minor et al., 2006). The structure was determined by molecular replacement using the H1N1 NP monomer as the search model. All calculations were done using the CCP4 suite (Collaborative Computational Project, Number 4, 1994), and model building was done in Coot (Emsley et al., 2004) and CNS (Brunger et al., 1998). Figures of protein structures were prepared with the program PyMOL (DeLano, 2002).

2.2.10 In vitro Pull-Down Assay

Purified wild-type PB2 '627-domain' was covalently immobilized on an NHS column, according to the manufacturer's instructions (GE Healthcare). Wild-type NP in PBS was then applied to the column and incubated for 1 h at room temperature. The column was washed with at least 20 column volumes of PBS. Bound protein was eluted with 1.5 M NaCl in PBS and analyzed by SDS-PAGE.

2.2.11 Surface Plasmon Resonance of NP-RNA Interaction

Biotinylated 2'-*O*-methylated RNA oligonucleotide with the sequence 5' UUU GUU ACA CAC ACA CAC GCU GUG 3' was immobilized on an SA sensor chip (GE Healthcare) until the surface density reached 30-35 response units (RU), according to the manufacturer's instructions (GE Healthcare). NP variants in different concentrations were allowed to pass through the chip surface. SPR measurements were carried out with a BIAcore 3000 at 25°C. Data were analyzed with BIAevaluation v. 4.1, for the calculations of association constant, dissociation

constant and the affinity of the interaction.

2.2.12 Surface Plasmon Resonance of NP-PB2 '627-domain' Interaction

Purified wild-type, [E627K,R630G] and R630G H5 PB2 '627-domain' variants were individually immobilized on a CM5 sensor chip (GE Healthcare) using an amine coupling kit until the surface density reached 700-800 RU. Wild-type and R150A NP variants in a series of concentrations were allowed to flow through the flow cells. A 2'-*O*-methylated RNA oligonucleotide with the above-mentioned sequence was also mixed with the protein variants in some experiment to assess the effect of RNA to the NP-PB2 '627-domain' interaction. SPR measurements were carried out with BIAcore 3000 at 25°C and data were analyzed with BIAevaluation v. 4.1.

2.2.13 Co-Immunoprecipitation of NP-NP

1 ug each of untagged and myc-tagged NP plasmids were transfected into 10^6 human kidney 293T cells in suspension. Co-immunoprecipitation was performed 48-hours post-transfection. Cells were resuspended in 50 mM Tris-HCl (pH 7.6), 150 mM NaCl, 1 mM EDTA, 1 % Triton X-100 (co-IP buffer) and lysed by sonication. The lysate was centrifuged at 16000 g for 10 minutes at 4°C. 150 U RNaseA (Sigma) was added into the supernatant, and it was incubated at 4°C overnight with or without anti-myc antibody. The mixture was then incubated with protein-A beads for 1.5 hours at 4°C with shaking. The beads were centrifuged and washed with co-IP buffer four times before being boiled in SDS-loading dye and analyzed by western blotting.

2.2.14 Co-Immunoprecipitation of RNP Complex

Either myc-tagged PA or myc-tagged PB2 were co-transfected with the other untagged polymerase subunits and NP. A similar protocol as described in section 2.2.12 was then followed, except that RNaseA was not added.

2.2.15 Influenza RNA Analysis by Primer Extension Assay

Human kidney 293T cells were used to reconstitute RNP complexes. 1.0 μg of each of the pcDNA-PB1, pcDNA-PB2, pcDNA-PA, pcDNA-NP, and pPOLI-NA-RT plasmids were diluted to a total volume of 250 μl in OptiMEM (Invitrogen) and subsequently added to a mix of 7.5 μl of Lipofectamine 2000 (Invitrogen) in 250 μl OptiMEM. The transfection mixture was incubated for 30 minutes before adding to 1.5 ml (about 10^6 cells) 293T cells in a suspension of minimal essential medium (MEM) containing 10% fetal calf serum in 35-mm dishes. Cells were harvested 48 hours post-transfection, and total RNA was extracted by TRIzol reagent (Invitrogen).

Primer extension assays were performed as described previously (Fodor et al., 2002; Vreede et al., 2004). Briefly, an excess of DNA primer (about 10^5 cpm), labeled at its 5' end with ^{32}P, was mixed with 5 μg of total RNA in 5 μl of water, and denatured at 95°C for 3 minutes. The mixture was cooled on ice and subsequently incubated at 45°C for 1 hour with the addition of 50 U SuperScript II RNase H^- reverse transcriptase (Invitrogen) in First Strand Buffer (Invitrogen). Two NA gene-specific primers and one 5S ribosomal RNA primer (used as an internal control) were used in the same reverse transcription reaction: 5'-GGACTAGTGGGAGCATCAT-3' (to detect vRNA), 5'-TCCAGTATGGTTTTGATTTCCG-3' (to detect mRNA and cRNA) and 5'-TCCCAGGCGGTCTCCCATCC-3' (to detect 5S rRNA). Reactions were stopped by the addition of 8 μl 90 % formamide and heating at 95°C for 3 minutes.

Transcription products were analyzed on 6 % polyacrylamide gels containing 7 M urea in TBE buffer and detected by autoradiography. Phosphorimage analysis by ImageQuant TL (GE Healthcare, Waukesha, WI, USA) was used for quantification. An unpaired Student's *t*-test was used for analysis of significance.

2.2.16 Influenza Polymerase Activity Analysis by Luciferase Assay

0.125 µg of each of the pcDNA-PB1, pcDNA-PB2, pcDNA-PA, pcDNA-NP, and pPOLI-Luc-RT plasmids and 0.0625 µg of pEGFP plasmid were diluted to a total volume of 12.5 µl in OptiMEM (Invitrogen) and subsequently added to a mix of 1.05 µl of Lipofectamine 2000 (Invitrogen) in 12.5 µl OptiMEM. The transfection mixture was incubated for 30 minutes in the 96-well plate before 75 µl (about 10^5 cells) 293T cells in minimal essential medium (MEM) containing 10% fetal calf serum was added into the well. GFP fluorescent signal was first measured 48 hours post-transfection. Afterwards, cells were lysed by Steady-Glo assay reagent (Promega, Madison, WI, USA) for 5 minutes, and the luminescence signal was measured. The polymerase activity was reported as the ratio of GFP signal to luminescence signal. An unpaired Student's *t*-test was used for analysis of significance.

2.2.17 Molecular Modeling of ring-like NP

After assembling nine NP monomers into a ring according to the EM structure of a mini-RNP (Martin-Benito et al., 2001), the linker regions were removed, and the tail loop was bent by ~90° toward the body domain. The linker regions were then rebuilt using the program *ArchPRED* (Fernandez-Fuentes et al., 2006).

2.2.18 Molecular Docking of NP-PB2 '627-domain' Interaction

The monomeric atomic coordinates of NP (PDB ID: 2Q06) and PB2 '627-domain' (PDB ID: 2VY8) were submitted to ClusPro (Comeau et al., 2004a; Comeau et al., 2004b) and for rigid protein-protein docking (http://cluspro.bu.edu/). The program was executed with default parameters. The two PDB files were also submitted to GRAMM-X (Tovchigrechko and Vakser, 2006), with the potential interacting surfaces roughly assigned (http://vakser.bioinformatics.ku.edu/resources/gramm/grammx/).

Chapter 3
Structural Study of Influenza A Virus Nucleoprotein

3.1 Introduction

Like the rabies, measles, and vesicular stomatitis viruses (VSV), the influenza A virus is a single-stranded, negative-sense RNA virus (NSRV). Its genome comprises 8 segments of RNA (vRNA) encoding 11 identified polypeptides. The two strain-defining surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA), are currently the main vaccine targets (Epstein et al., 2005). Both HA and NA exist in several different forms and have high mutation rates. The antigenic drift allows the influenza A virus to escape from established humoral immunity in humans (Shu et al., 1993). Vaccines must then be reformulated, specifically for serologically distinct viruses in each influenza season. The time delay in large-scale vaccine production against a suddenly appearing, new pandemic strain could cause a global disaster. By contrast, internal viral proteins, such as the nucleoprotein (NP), are substantially more conserved, and hence are ideal targets for T-cell-mediated immunity (Gschoesser et al., 2002).

Similar to other NSRVs, the genome of the influenza virus is encapsidated by NP. The primary function of NP is to condense the segmented genomic RNA into a helical nucleocapsid and, together with three polymerase subunits, PA, PB1, and PB2, to form a ribonucleoprotein (RNP) particle for RNA transcription, replication,

and packaging. The RNP, rather than the naked vRNA, is the template for transcription and replication (Bishop et al., 1971). However, NP is not only a structural protein for RNA binding, it also acts as a multifunctional key adaptor molecule for interactions between virus and host cell (reviewed in Portela and Digard, 2002). NP is likely to be the major switching factor that determines whether genomic vRNA is transcribed into mRNA encoding viral proteins or used as template to synthesize cRNA for genome replication (Beaton and Krug, 1986). NP also interacts with a number of viral and cellular proteins. In addition to encapsidating viral RNA (Kobayashi et al., 1994), NP also forms homo-oligomers to maintain the RNP structure (Prokudina-Kantorovich et al., 1996) and binds to the PB1 and PB2 subunits of the RNA polymerase (Biswas et al., 1998). Cellular proteins with which NP interacts include importin, F-actin, and CRM1/exportin-1 (reviewed in Portela and Digard, 2002). These interactions mediate RNP trafficking in and out of the nucleus and are thus vital for viral propagation and assembly.

Its multiple functions and conserved protein sequence make NP an excellent drug target for all influenza A virus subtypes. In this context, an atomic resolution structure of NP is of crucial importance. So far, electron microscopy (EM) has produced a very low resolution structure of the mini-RNP (Martin-Benito et al., 2001; Area et al., 2004), and more recently, X-ray crystallography has yielded a 3.2-Å resolution structure of a homologous NP from human origin (A/WSN/1933; H1N1) (Ye et al., 2006). In the current study, we have determined the 3.3-Å crystal structure of A/HK/483/97 (H5N1) NP (PDB code: 2Q06) from avian origin, allowing the resolution of some missing elements in H1N1 NP structure and providing explanations for the different trimerization tactics in NP.

Owing to the establishment of protein sequence database and the extensive sequencing studies of influenza A viruses, a large number of cDNA sequences of various influenza virus proteins are available. So far, there are more than 2,500 partial or complete NP protein sequences available in the NCBI database. Multiple sequence alignment of these 2,500 sequences would allow the identification of amino acid polymorphism in NP. A cut off value of 0.5 % as in single nucleotide polymorphism (Zhang and Sun, 2005) is employed to distinguish whether the protein sequence variations are the results of polymorphism or simply random mutations and sequencing errors. Residues with a single amino acid over 12 counts (out of 2500 NP sequences) are defined as conserved while residues with more than one amino acid variations are polymorphic. Residues which can bear four or more variations are hypervariable. Together with the atomic structure of NP, we can have a better understanding of its structure-function relationship.

3.2 Results and Discussion

3.2.1 Biophysical Characterization of NP

H5N1 NP was expressed as a maltose-binding protein (MBP)-fusion protein in *E. coli* and purified to 95% homogeneity (Figure 3.1). Electron microscopy (EM), done by our collaborator Zongli Li and Thomas Walz, of negatively stained samples showed that, after tag removal and RNase treatment, NP exists in solution as a mixture of trimers and tetramers (Figure 3.2A). In the presence of RNA, NP assembles into higher-order oligomers (Figure 3.2B). Light scattering studies of RNase-treated NP in solution suggested a molecular mass of around 193 kDa (Figure 3.3), also corresponding to a mixture of trimers and tetramers, but only trimers were incorporated into the 3-dimensional crystals.

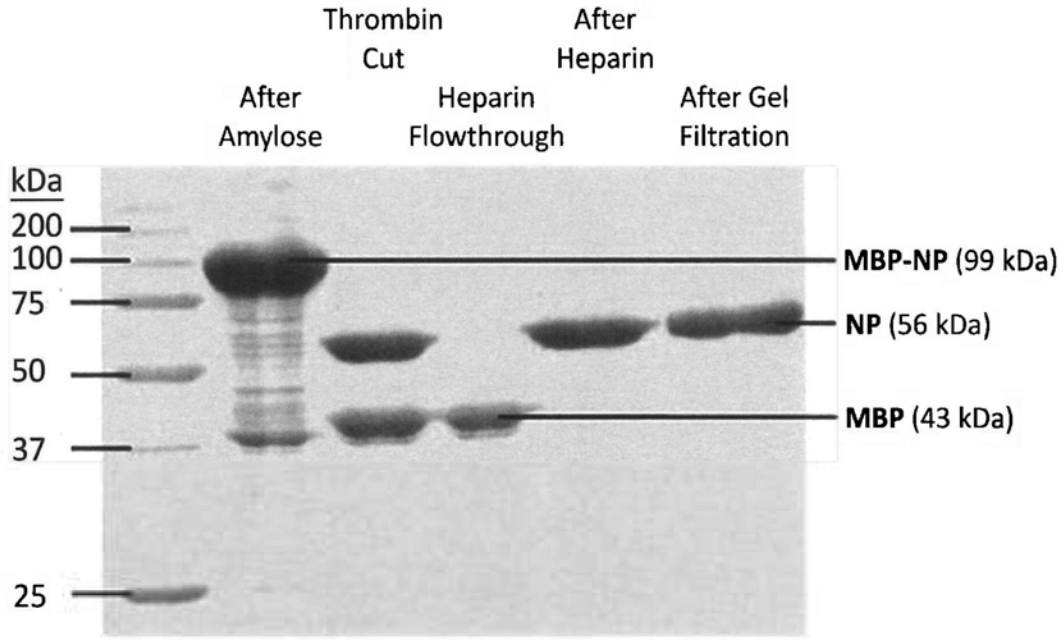


Figure 3.1 Purification of NP. MBP-NP was first captured by amylose column; the MBP-tag was then cleaved by thrombin and RNA removed by RNaseA. NP was then purified by heparin column and gel filtration column; and attained a homogeneity of more than 95 %.

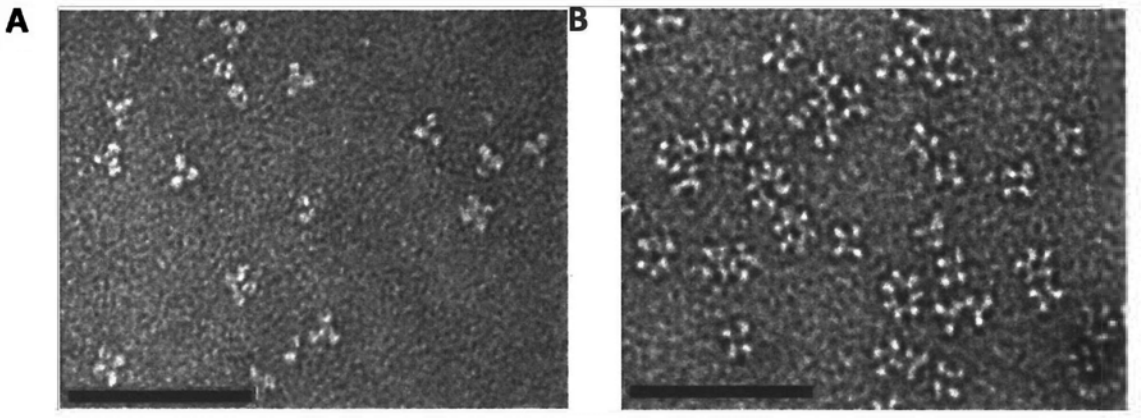


Figure 3.2 Electron microscopy of NP in the presence and absence of RNA. (A) RNase-treated NP existed in solution in a mixture of trimers and tetramers. **(B)** Higher-order oligomers were observed when a 51-nt RNA was added to an RNase-treated NP sample. Scale bars = 50 nm. Experiments were done by Zongli Li.

3.2.2 Structural Features of NP

3.2.2.1 Atomic structure of NP

H5N1 NP formed small tetrahedral crystals (Figure 3.4) with the cubic space group $P2_13$. Crystals were rapidly frozen after stepwise soaking them in increasing concentrations of sucrose up to a concentration of 20% in the crystallization buffer. A 3.3-Å resolution data set from a single, frozen crystal was collected at 100 K at the 19ID beamline (with X-rays at a wavelength of 0.97845 Å) of the Structural Biology Center at the Advanced Photon Source of Argonne National Laboratory using the program SBCcollect. The data were processed and scaled with the HKL3000 suite (Minor et al., 2006). The crystal belonged to space group $P2_13$ and the structure of H5N1 NP was determined by molecular replacement, using the H1N1 NP monomer (PDB code: 2IQH, Chain A) (Ye et al., 2006) as the search model (its partial tail loop was not included in the calculations). One crystallographic asymmetric unit comprises two H5N1 NP molecules, referred to as molecules A and B. Even in the initial 2Fo-Fc map, the entire tail loops of both molecules were clearly defined by continuous electron density at a 1-sigma contour level. The two linkers as well as some residues that were missing from the C-terminal region of the H1N1 model were, therefore, built based on the difference maps. The structure was refined to R/Rfree = 20.2%/27.9% (Table 3.1). Ramachandran plot statistics show that all residues were found in most favored/ allowed regions. Overall, 467 amino acids were modeled for molecule A (residues 22-78 and 87-496), and 465 amino acids were modeled for molecule B (residues 21-77 and 88-495) (Figure 3.5A).

The three-dimensional structure NP is organized into the head domain, the body domain and the tail loop/linker region. The entire polypeptide chain goes back and forth between the head and body domains several times. The structure

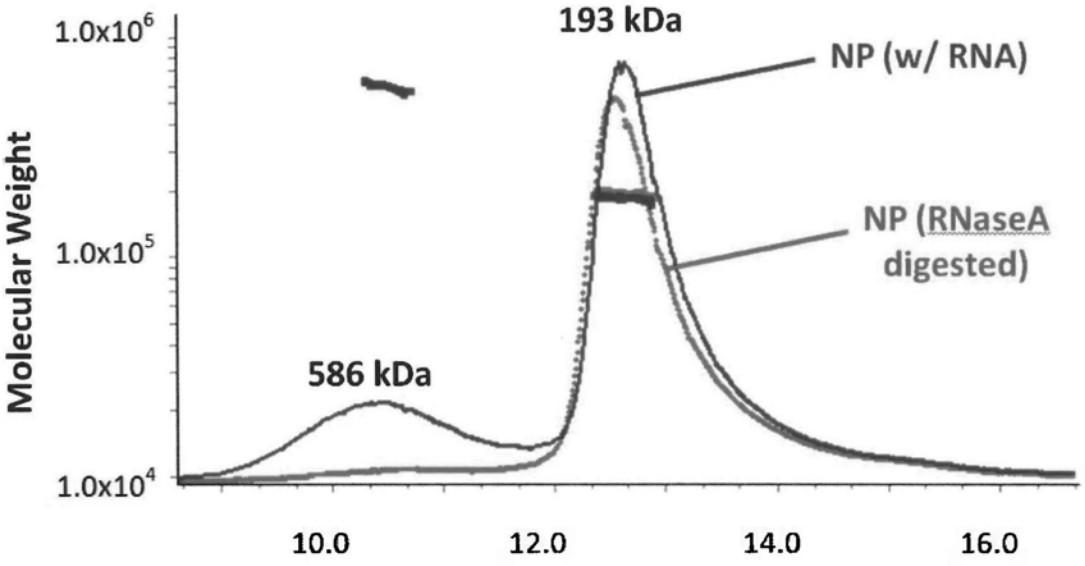


Figure 3.3 Static Light scattering of untreated and RNase-treated NP. Two distinct populations were found in an NP sample not treated with RNase (blue). The peak at 586 kDa corresponds to higher-order oligomers; the peak at 193 kDa corresponds to the mixture of trimers and tetramers. After treatment with RNase, the high molecular mass population disappeared (red). The RNase-treated sample was less polydispersed (0.5%) than the untreated sample (2%) and therefore more suitable for crystallization.

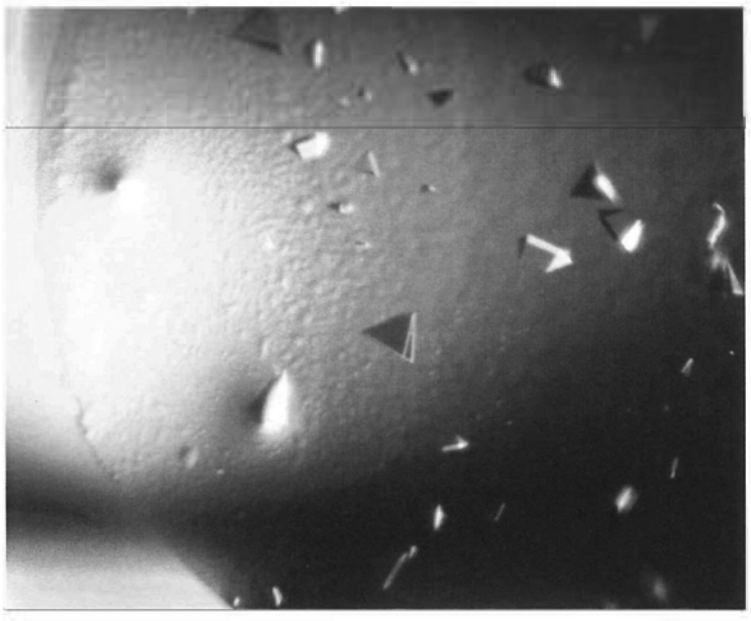


Figure 3.4 Protein crystals of NP. Protein crystals appeared in 0.15 M KCl, 0.01 M $MgCl_2$, 0.1 M sodium cacodylate, pH 6.2, and 7% PEG 8000, after optimization.

Table 3.1 Diffraction data and refinement statistics of NP crystals.

Diffraction Data		
Wavelength (Å)		0.97845
Resolution (Å)*		48.564-3.301 (3.38-3.30)
Cell constant (Å)	*a*	153.578
	b	153.578
	c	153.578
Unique reflections		18405
Redundancy		**5.0**
Mean I/σ (I)*		11.16 (2.1)
R_{merge} (%)*		14.1 (66.0)
Completeless (%)*		99.7 (99.1)
Refinement Statistics		
Resolution		46.32-3.30
R_{work}/R_{free}		0.202/0.279
# of R_{free} reflections		945 (5.1 %)
# of protein atoms		7343 (932 residues)
Average B-factors		93.109 $Å^2$
R.m.s. deviations	Bond	0.009 Å
	Angle	1.236°
Ramachandran Plot Statistics		
Most favored regions		656 (80.2 %)
Additional allowed regions		146 (17.8 %)
Generously allowed regions		16 (2.0 %)
Disallowed regions		0 (0.0 %)

* Data in brackets are from the high resolution shell.

begins at the body domain with helices 1 to 4 (residues 1–148), followed by the first segment of the head domain (residues 150–263), encompassing helices 5 to 11. Then a loop region links the head domain back to the second segment of the body domain (residues 277–396), comprising helices 12 to 17, after which comes the tail loop/linker region (residues 397–436), which extends out to insert into the neighboring NP molecule. The tail loop/linker region is subdivided into 'Linker 1' (residues 397–401), tail loop (residues 402–428) and 'Linker 2' (residues 429–436). Then, helix 18 in the tail loop interacts with helix 19 in the second segment of the head domain (residues 437–451). Connected by another loop region, the chain returns to the third segment of the body domain (residues 464–498) at the C-terminus (Figure 3.5C).

The crystal structure of H5N1 NP is very similar to H1N1 NP (Ye et al., 2006). The two NP sequences share 94% sequence identity, and after aligning 398 residues (residues 22–72, 92–202, 213–396, 438–489) the root-mean-square (RMS) deviation between the two crystal structures is 1.0 Å. Our H5N1 NP structure reveals, however, several novel and important features that have not previously been seen in the H1N1 NP structure. First, our map of H5N1 NP shows well-defined electron densities for residues 397–401 and 429–437 (Figure 3.5B), which were missing in the H1N1 NP model and now connect the extended tail loop (residues 402–428) to the main body. Second, seven additional C-terminal residues (residues 490–496) could be modeled in our H5N1 NP structure (Figure 3.5B) and were found to form coils and bends. These residues are located near the insertion site of the tail loop and are thought to play a regulatory role in NP oligomerization (Elton et al., 1999). Third, more residues (residues 72–78 and 87–92) could be built in a flexible, basic loop toward the concave region of the molecule postulated to be an RNA-binding

groove (Ye et al., 2009). Finally, H5N1 and H1N1 show very interesting differences in the way they trimerize, as will be discussed in detail below.

H5N1 NP contains five cysteine residues. Remarkably, none of them form disulfide bonds with each other. Cys-44, Cys-164, and Cys-223 form hydrogen bonds with the side chains of other amino acids. Interestingly, the C-α atoms of Cys-275 and Cys-333 are only 5.6 Å apart, well within the range of disulfide bond formation, but their side chains point away from each other. A previous study suggested the formation of a transient disulfide bond during the conformational maturation of NP (Prokudina et al., 2004), and Cys residues 275 and 333 would be very good candidates to play a role in this process.

3.2.2.2 Conservation of NP secondary structure

Altogether, NP has 19 α-helices (242 residues, 48.6%) and 8 β strands (24 residues, 4.8%), while the rest are loops (232 residues, 46.6%) (Figure 3.5C). NP is therefore highly helical. On the other hand, the longest strands are only composed of 4 residues. After comparing more than 2,500 NP protein sequences, it is found that helices are the most conserved structural element, with 72.7% sequence conservation. This is followed by loops (68.1%) and strands (58.3%).

When examining the helical elements of NP (Table 3.2), some are found to be strikingly conserved. The sequences of helices 6 and 7 are completely conserved among all the available NP sequences. Several other long helices, including helices 4, 9 and 10, are also more than 80% conserved. It would be interesting to find if these helices may play a role in the folding of NP or maintaining the structure. With the available H5N1 and H1N1 NP structures, the roles of the different helices are

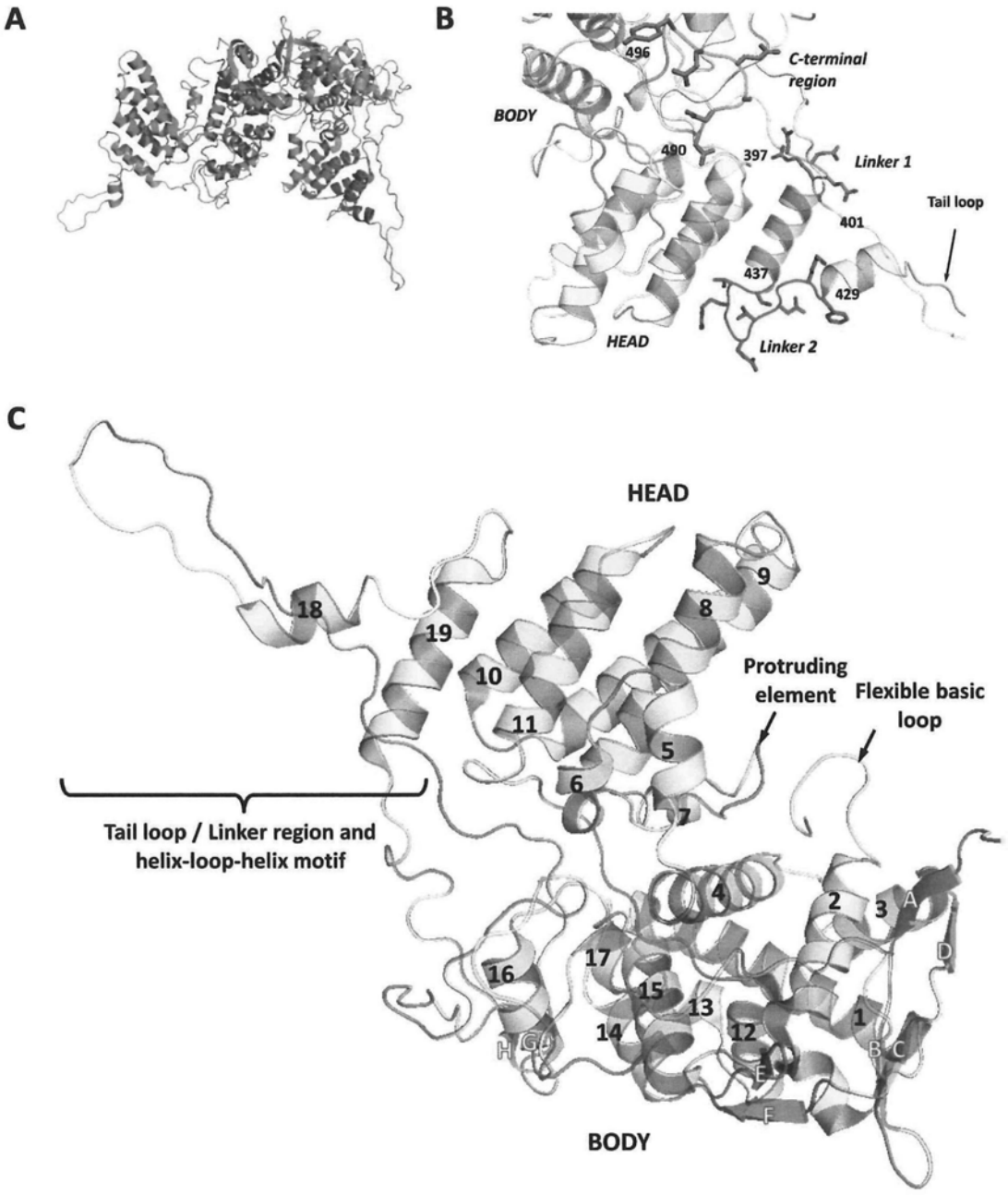


Figure 3.5 Crystal structure and domain organization of NP. **(A)** Each asymmetric unit consists of two NP molecules in the crystal structure. **(B)** Three segments of H5N1 NP could be modeled that were not visible in the H1N1 NP structure. Residues 397-401 (green) and 429-437 (purple) are linkers that connect the tail loop to the main body of NP. C-terminal residues 490-496 (cyan) are close to the tail loop insertion site and may contribute to the regulation of NP oligomerization. **(C)** NP is organized into the head domain, body domain and the tail loop/linker region. The protruding element and the flexible basic loop are located at the arginine-rich groove between the head the body domains. All the strands (labled as A to H) are located at the body domain.

Table 3.2 Sequence conservation and functions of helices

Domain	Helix	Residues	Percentage Conservation[1]	Functions	
				Interpretation from deletion mutants[2]	Interpretation from structure
Body	1	23-47	72.0	RNA binding, PB2 binding	
	2	50-69	60.0	RNA binding, PB2 binding	Connecting to the flexible basic loop
	3	113-123	54.6	RNA binding, PB2 binding	
	4	130-148	84.2	RNA binding, PB2 binding	
Head	5	150-157	87.5	RNA binding, PB2 binding	RNA binding
	6	160-167	100.0	RNA binding	Interaction with the tail loop
					Connecting to the protruding element
	7	177-182	100.0	NP-NP homo-oligomerization	Connecting to the protruding element
	8	185-202	61.1	NP-NP homo-oligomerization	
	9	210-230	85.7	NP-NP homo-oligomerization	
	10	232-244	84.6	NP-NP homo-oligomerization	
	11	249-263	80.0	NP-NP homo-oligomerization, PB2 binding	Trimerization interface
Body	12	277-287	54.6	NP-NP homo-oligomerization, PB2 binding	
	13	290-295	83.3	NP-NP homo-oligomerization, PB2 binding	
	14	301-309	66.7	NP-NP homo-oligomerization, PB2 binding	
	15	321-334	78.6	NP-NP homo-oligomerization, PB2 binding	Interaction with the tail loop
	16	340-349	60.0	NP-NP homo-oligomerization, PB2 binding	Interaction with the tail loop
	17	354-358	80.0	NP-NP homo-oligomerization, PB2 binding	
Tail loop	18	421-428	50.0	NP-NP homo-oligomerization, PB2 binding	Helix-loop-helix motif for trimerization
Head	19	437-451	60.0	NP-NP homo-oligomerization, PB2 binding	Helix-loop-helix motif for trimerization

[1]Percentage conservation is calculated by (number of non-polymorphic residues / number of residues in the helix * 100 %)

[2]Prokudina-Kantorovich et al., 1996; Skorko et al., 1991; Elton et al., 1999; Albo et al., 1995.

better defined (Table 3.2).

In influenza A NP the head domain has an exceptionally high helical content of 80.6%, which is much higher than the 48.6% of the overall structure (Table 3.3). There are no β-strand structures in the head domain. The head domain of NP is highly conserved among influenza A virus NP. 80.7% (104 out of 129) of the residues in the head domain are not polymorphic, in comparison to the 68.9% of the overall structure. When the NP structures of rabies virus (Albertini et al., 2006) and VSV (Green et al., 2006) are compared, it is found that these two NP are also organized into head and body domains and the tail loop region. Interestingly, the head domains of rabies and VSV NP are also found to be more structurally conserved than the body domain, as they share similar arrangement of secondary element (Luo et al., 2007).

The body domain is less helical than the head domain, as it only contains 42.9% of helical content (Table 3.3). Although the strands are scattered throughout the sequence of NP, all of them are located at the body domain and being exposed to the solvent.

Out of the 12 hypervariable residues in NP, 8 are found in the body domain, 3 in the tail loop/linker region while only 1 is found in the head domain, which again demonstrates the conserved nature of the head domain. Structurally speaking, 7 of the 8 body domain hypervariable residues (D53, R100, D101, V371, A373, D375, S377) are localized at the very end of the domain, and substantially exposed to solvent. These residues are likely to provide the flexibility in interacting with various partners, such as polymerase subunits, in the different strains of the

influenza virus.

3.2.2.3 A comparison of the accessible surface area between monomeric and trimeric NP

With the structure in hand, we have calculated the accessible surface area (ASA) of the protein molecule and of individual residues by the program VADAR (Willard et al., 2003). The calculated ASA (in $Å^2$) is the exposed surface area of the residues in which a water molecule can gain access to. The fractional ASA evaluates the exposed or hidden nature of a particular residue when comparing with the standard values. The main chain and side chain fractional ASA are usually within minor differences in most cases. Since protein contacts are mainly contributed by the side chain atoms (Faure et al., 2008), residues with their side chain fractional ASA greater than 50% are considered exposed to the solvent.

Assuming the monomeric NP and trimeric NP have similar conformation (Janin et al., 2008), comparing the ASA between the monomer and trimer would shed light on the molecular contacts required to form trimer, and even oligomer. Residues with more than 50% decrease of side chain ASA in trimeric NP are first identified. These residues are significantly more conserved. Out of 78 residues, only 15 display polymorphic nature and only 1 of them is hypervariable among the NP sequences. The 78 residues are mainly concentrated in five regions at residues 161–166, 250–272, 330–347, 387–463 and 486–490 (Figures 3.6A and 3.6B). The region at residues 387–463, which contains the tail loop, displays the most dramatic decrease in the ASA. The other four regions contain the trimerization interface and also make up of the groove for the tail loop to insert. All the five regions are located at the C-terminal of NP, which is consistent with the previous findings that

Table 3.3 Domain characterization of NP

Domain	Residues	Number of Conserved Residues	Number of Polymorphic Residues	Number of Hypervariable Residues	Sequence Conservation	Percentage helix	Percentage strand	Percentage loop
Head	150-263	95	19	1				
	437-451	9	6	0				
					80.7	**80.6**	**0**	**19.4**
Body	1-148	100	48	3				
	277-396	78	42	5				
	464-498	22	13	0				
					66.1	42.9	7.9	49.2
Tail loop	397-401	2	3	0				
/ Linker	402-428	19	8	1				
regions	429-436	6	2	2				
					67.5	20.0	0	80.0

Note: Sequence conservation is analyzed from around 2500 NP sequence in NCBI database. Secondary structure is analyzed from the H5N1 NP structure.

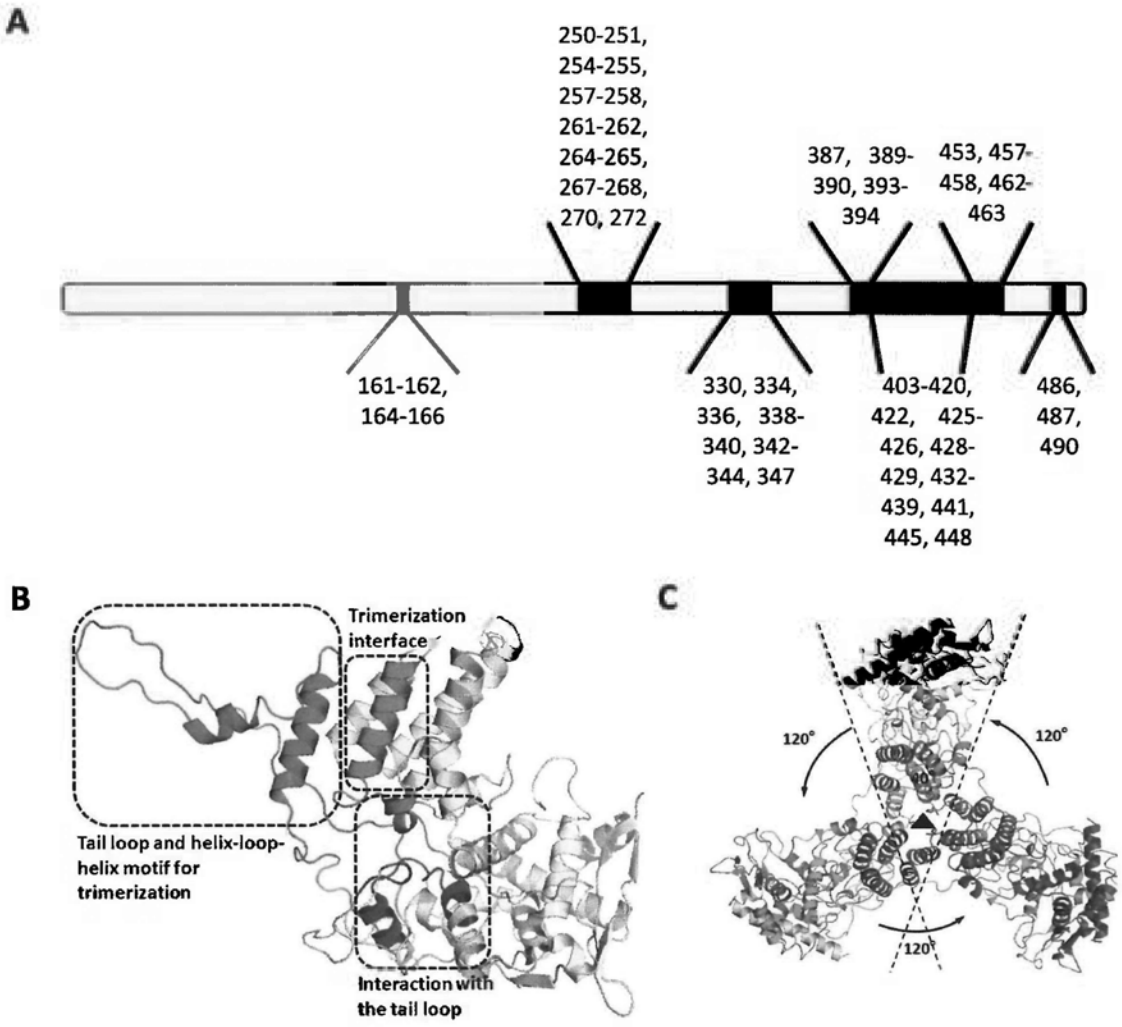


Figure 3.6 Trimerization of NP. **(A)** Regions which are more hidden in trimeric NP structure are shown with colors. Residues with more than 50% decrease in the side chain accessible surface area are labeled. **(B)** Purple region (residues 387–463) contains the helix-loop-helix motif and the tail loop/linker region, which are essential for trimerization. The green region (residues 250–272) provides molecular contact for trimerization. These two regions may have more extensive contacts with their neighboring NP protomer during the oligomeric state. Blue (residues 330–347) and cyan (residues 161–166) regions are the groove for tail loop insertion. The red region (residues 486–490) may inhibit the NP homo-oligomerization. **(C)** The crystallographic trimer of NP.

residues 189–358 and 371–465 are important for NP self association (Elton et al., 1999).

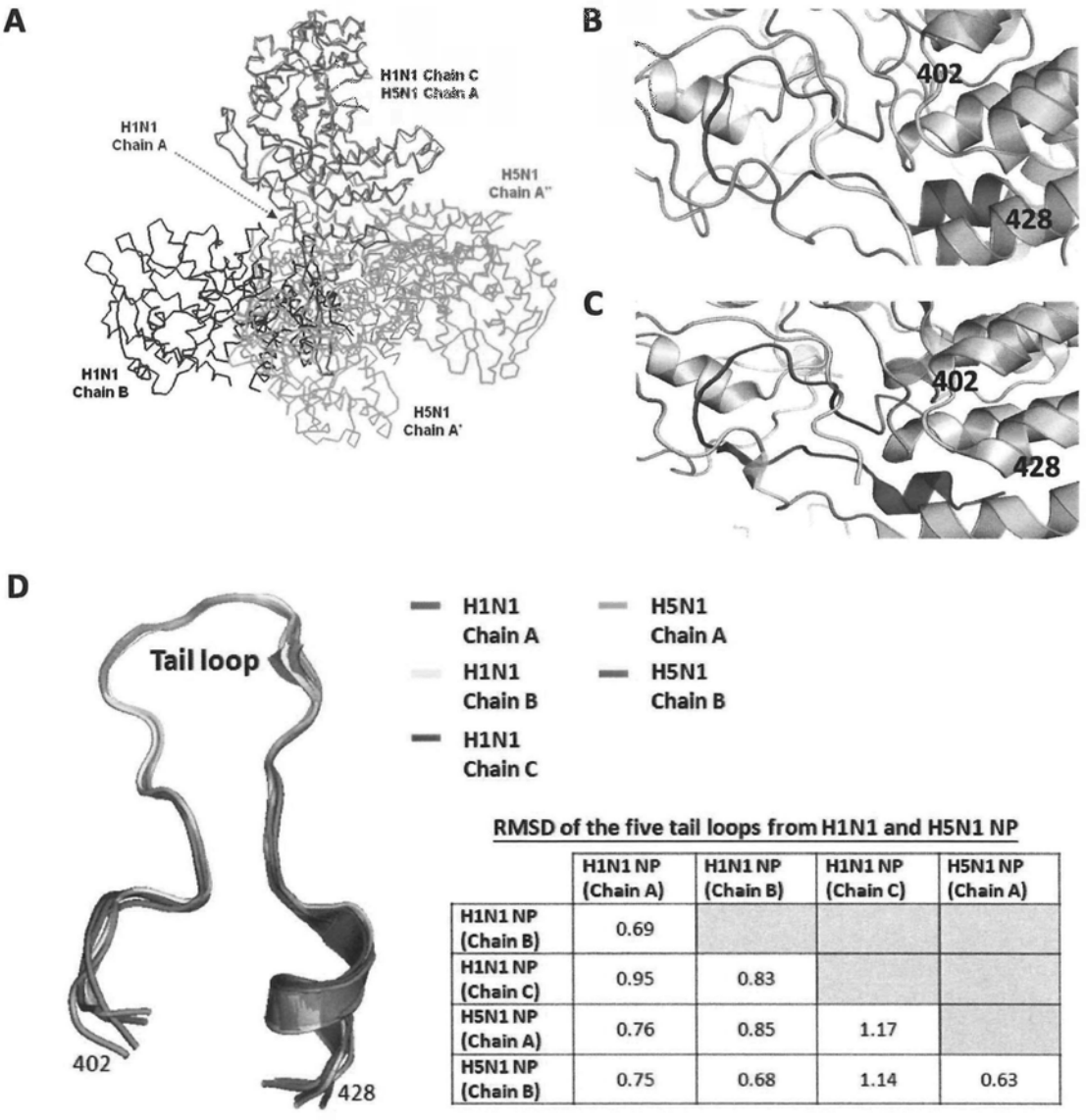
3.2.3 Tail Loop Insertion

In order to maintain a stable RNP structure, NP forms homo-oligomers to wrap around RNA. The affinity (K_D) of NP-NP interaction is 200 nM (Elton et al., 1999; Prokudina-Kantorovich et al., 1996).

In our H5N1 NP crystal, each of the two molecules in the asymmetric unit is part of a trimer about the crystallographic 3-fold axis, so that the intermolecular interactions within each trimer are identical (Figure 3.6C). By contrast, in the H1N1 crystal, the asymmetric unit contains a single trimer with the three protomers forming slightly different interactions with one another. Although the trimers formed by H5N1 NP and H1N1 NP are very different and cannot be superimposed (Figure 3.7A), the subunits in both trimers are connected by the same principle: the insertion of the long tail loop (residues 402-428) from one NP protomer into the body domain of a neighboring protomer (Table 3.4; Figure 3.7B and 3.7C).

There are currently five tail loop structures available, two from H5N1 NP (Ng et al., 2008) and three from H1N1 NP (Ye et al., 2006). Structural alignment of these tail loops (Figure 3.7D) shows that the conformation of the loop is much conserved, with root-mean-square-deviation (RMSD) between 0.63 Å and 1.17 Å.

Based on the structure of NP, we have recently mapped the crucial residues on the tail loop and the body domain of NP which mediate its homo-oligomerization (Chan et al., 2010) (Section 1.4.1).



RMSD of the five tail loops from H1N1 and H5N1 NP

	H1N1 NP (Chain A)	H1N1 NP (Chain B)	H1N1 NP (Chain C)	H5N1 NP (Chain A)
H1N1 NP (Chain B)	0.69			
H1N1 NP (Chain C)	0.95	0.83		
H5N1 NP (Chain A)	0.76	0.85	1.17	
H5N1 NP (Chain B)	0.75	0.68	1.14	0.63

Figure 3.7 H5N1 NP and H1N1 NP have identical interactions of the tail loops. **(A)** Superimposition of one molecule (chain A) of the H5N1 NP trimer (in color) with a molecule (chain C) of the H1N1 NP trimer (gray) shows that the other two molecules do not superimpose with each other. Nevertheless, the interaction of the tail loop of one NP molecule with the neighboring protomer in **(B)** H5N1 and **(C)** H1N1 is virtually identical. **(D)** Structural alignment and the RMSD of the five tail loop structures (residues 402–428) reveals its conservation in tertiary structure, which correlates with its conserved primary sequence.

Table 3.4 Inter-subunit interactions of NP in the crystallographic trimers of chains A and B

Chain A			Chain B		
S402	--	**R162**	S402	--	**R162**
G404	--	R162	S403	--	T151
Q405	--	Y487			D491
		F489	Q405	--	Y487
		F489			**F489**
I406	--	R162			F489
		Y487	I406	--	R162
S407	--	**L264**			Y487
		R267	S407	--	**S165**
		Y487			**L264**
I408	--	D340			L264
		R267			**R267**
Q409	--	**S165**			Y487
		F338	I408	--	**R267**
		V270			E339
P410	--	**H272**			D340
		F458	Q409	--	**S165**
T411	--	H272			V270
		S335			**F338**
		E339			F338
		T390	P410	--	**H272**
F412	--	**E339**			H272
		E339			E339
		T390			F458
S413	--	**I388**	T411	--	H272
		T390			**S335**
		R461			**E339**
V414	--	A387			E339
		F458			R389
		R461			**T390**
		P477	F412	--	**E339**
Q415	--	F458			E339
		Q459			T390
		Q459			Y387
		R461			R389
R416	--	**E339**	S413	--	**I388**
		V456			T390
		S457			R461
N417	--	E454	Q415	--	V476
L418	--	**R267**			F458
		R267			**Q459**
P419	--	D340			Q459
		S486			**R461**
F420	--	I265	R416	--	**E339**
		R452			E339
		P453			V456
R422	--	**E449**			**S457**
		R452	N417	--	E454
		P453			V456
I425	--	L264	L418	--	**R267**
		I265			R267
		M448	P419	--	**R267**
M426	--	E449			D340
A428	--	R261			S486
			F420	--	I265
					R452
					P453
			D421	--	**Y487**
			R422	--	E449
					R452
					P453
			I425	--	L264
					I265
					E449
			A428	--	R261

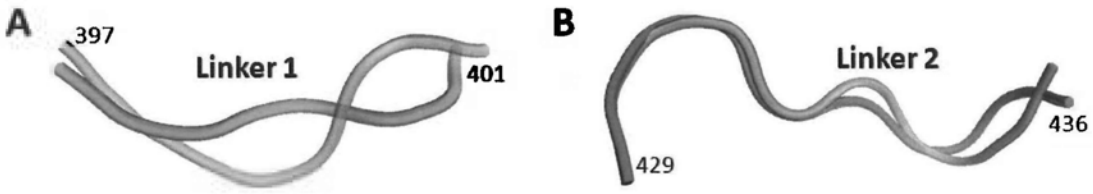
Note: The interaction in bold are hydrogen bonds while the rest are van der Waal's interaction.

3.2.4 Flexibility of Linkers Allows NP to Form Different Oligomers

In the H5N1 NP, the long loop protrudes straight from one protomer into the neighboring protomer, but forms an angle of 70° in the H1N1 NP. These very different conformations suggest that the two linkers (residues 397–401 and 429–437) have a high degree of flexibility. It is conceivable that this linker flexibility is important for converting NP trimers to higher-order oligomers during virus packaging.

The linker flexibility is shown by two pieces of evidence. First, the electron density of the linker regions in H1N1 NP structure is missing, suggesting its flexible conformation. Second, there are two linker structures resolved in H5N1 NP. When the two 'Linker 1' structures are aligned (Figure 3.8A), the backbone RMSD is 1.46 Å. This high RMSD shows its high flexibility. 'Linker 2' structures (Figure 3.8B) are more superimposable than those of 'Linker 1', but still slightly more flexible than that of the five tail loops. Third, the polymorphic nature of the linker regions also correlates to its flexibility. 'Linker 1' contains three polymorphic residues out of five while 'Linker 2' contains two hypervariable residues (Figure 3.9).

The formation of a trimer brings the three helix-loop-helix motifs near the C-terminus of each NP protomer into close contact (Figure 3.6C). Helix 18 of the helix-loop-helix motif is formed by residues 421-428 and helix 19 by residues 437-451. Since the trimers formed by H5N1 NP and H1N1 NP are different, the molecular contacts between the helix-loop-helix motifs are also different in the two trimers. This distinction can even be seen in the trimers formed by molecules A and B in our H5N1 NP structure. In the trimer formed by molecule A, helices from two adjacent protomers are held together by a salt bridge between R422 in helix 18 of one protomer and E449 in helix 19 from the adjacent protomer (Figure 3.10A).



RMSD of the two linkers in H5N1 NP

	H5N1 NP (Chain A)	
	Linker 1	Linker 2
H5N1 NP (Chain B)	1.46	0.99

Figure 3.8 Structural alignment of Linkers. **(A)** Structural alignment of Linker 1 (residues 397–401) from molecule A and B of H5N1 NP and its high RMSD suggest that this region is poorly aligned, which coincides with the polymorphic nature of the primary sequence. **(B)** Structural alignment of Linker 2 (residues 429–436) shows that it is better aligned than Linker 1

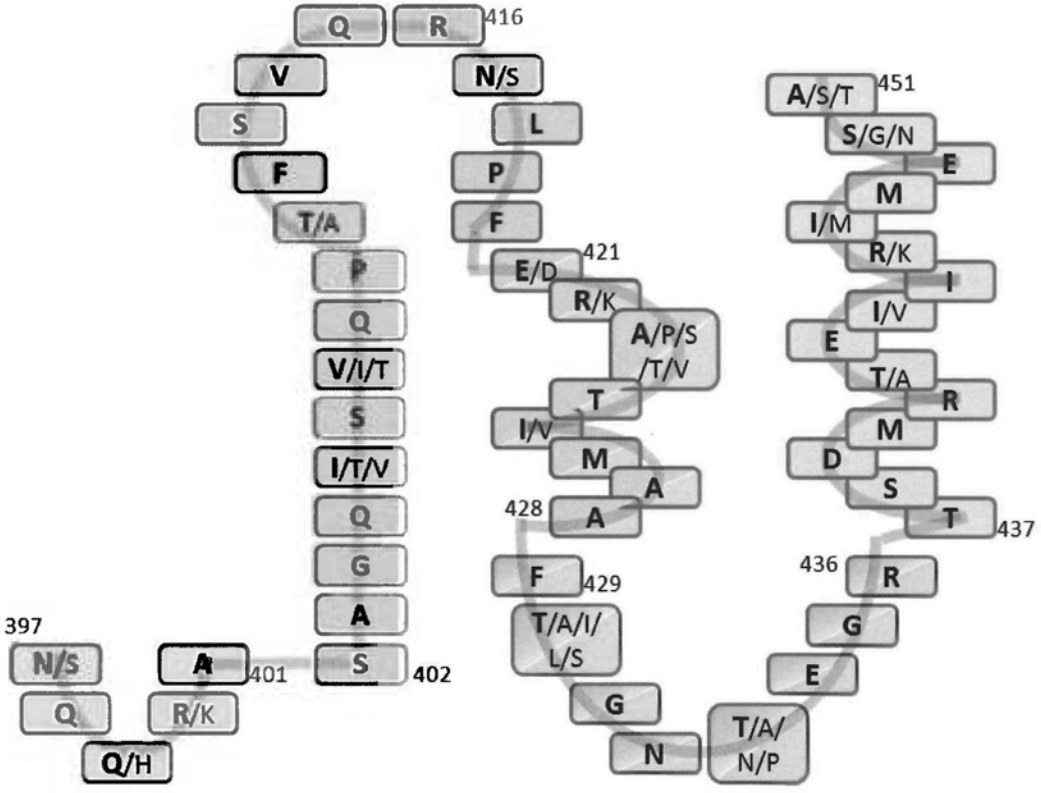


Figure 3.9 Sequence polymorphism of the tail loop/linker region and the helix-loop-helix motif. Linkers (residues 397–401, 429–436) (shown in green) are polymorphic while the tail loop (shown in blue) excluding helix 18 (residues 402–420) is conserved. Helix-loop-helix motif (residues 421–451) (shown in red) is polymorphic, with the presence of three hypervariable residues.

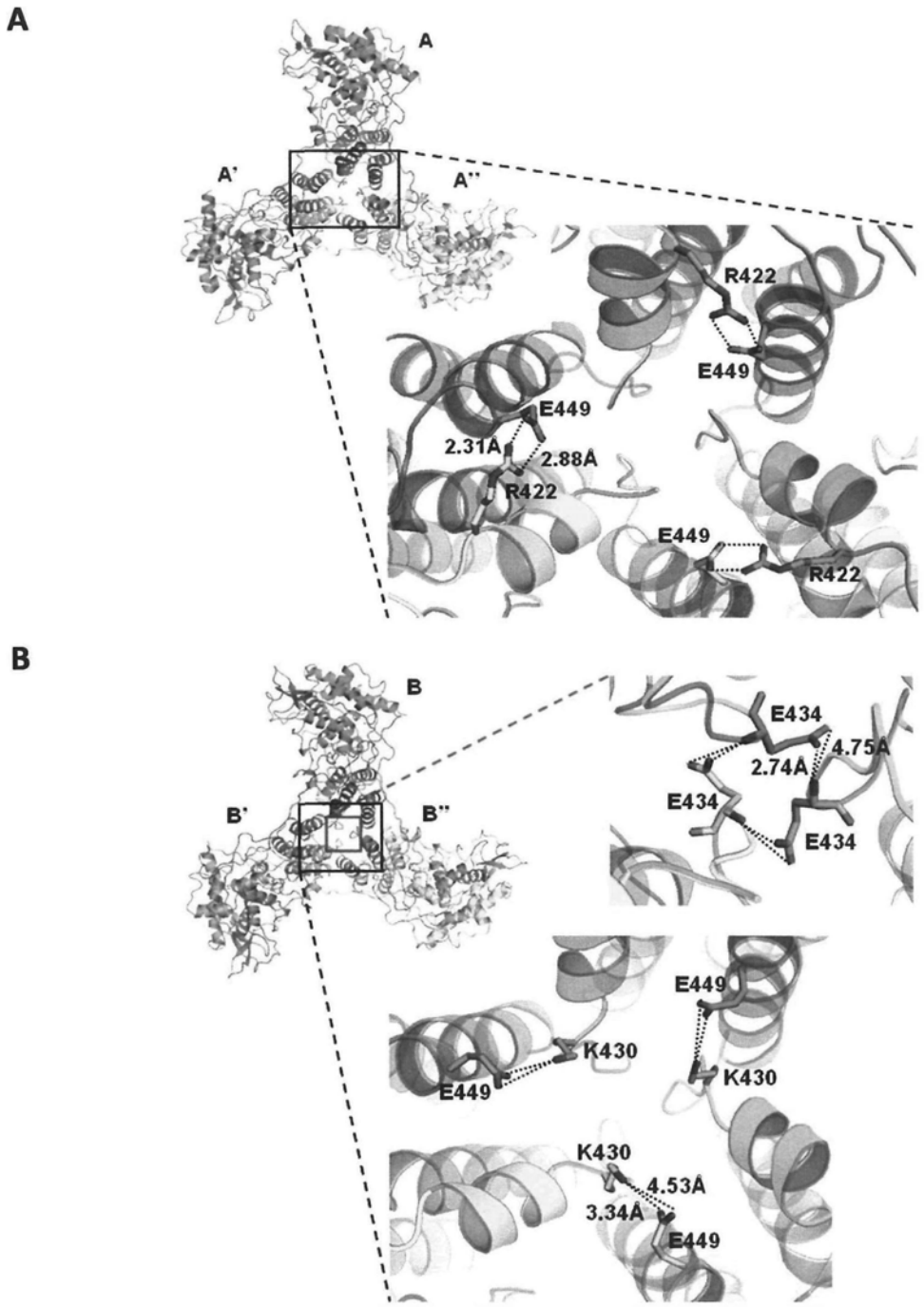


Figure 3.10 Molecules A and B of H5N1 NP have different tactics in trimer formation. **(A)** The trimer formed by H5N1 NP molecule A is held together by a salt bridge formed between R422 in helix 18 of the helix-loop-helix motif in one protomer with E449 in helix 19 of the helix-loop-helix in the neighboring protomer. **(B)** The trimer formed by H5N1 NP molecule B is held together by hydrogen bonds formed between the E434 residues of the three protomers (top panel). A salt bridge between residues K430 and E449 stabilizes the two helices of the helix-loop-helix motif in each protomer (bottom panel).

This interaction was not observed in the trimer formed by H1N1 NP. In the trimer formed by molecule B, the two helices of the helix-loop-helix motif are held close together by a salt bridge between K430 in helix 18 and E449 in helix 19. The helix-loop-helix motifs of neighboring protomers are linked through hydrogen bonds between the side chain of E434 in one protomer with the NH group of E434 in the neighboring protomer (Figure 3.10B). Although residues 429-437 were not resolved in the crystal structure, H1N1 NP has a threonine at position 430 instead of a lysine. Therefore, the salt bridge linking protomers in the H5N1 NP trimer is unlikely to be formed in the H1N1 NP trimer, offering a possible explanation for the different configurations of the two trimers. It is worth noting that a survey of more than 300 influenza virus A NP sequences showed that only 25 had a lysine residue at position 430, and 20 of these were from influenza A viruses of the H5N1 subtype. Whether this H5N1-specific substitution bears any significance for virus assembly remains to be determined.

Because the H5N1 NP and H1N1 NP trimers differ substantially (Figure 3.7A), it is unlikely that the trimer is the physiologically relevant RNA-binding unit in influenza A virus, as has previously been suggested (Ye et al., 2006). In physiological conditions, NP exists in the form of helical oligomer (Ruigrok and Baudin, 1995). Three dimensional reconstruction of a mini-RNP indicates that NP molecules are more closely packed during oligomerization (Martin-Benito et al., 2001). Therefore, we propose a model in which nine NP molecules organize into a ring-like structure (Figure 3.11A). This model is based on structures of nucleoprotein-RNA complexes of VSV (Green et al., 2006) and rabies virus (Albertini et al., 2006) and the low-resolution EM structure of a mini-RNP (Area et al., 2004; Martin-Benito et al., 2001). On the basis of our model, we can propose the surfaces of NP that may be

involved in its interaction with other viral and cellular components. In our model, NP assembles into a ring using the same tail loop as it uses to form a trimer, since mutant viruses in which the tail loop of NP (residues 418-426) was deleted could not be rescued (Berkhoff et al., 2006). To assemble nine protomers into a ring, the tail loop has to bend by ~90° toward its body domain, compared to its conformation in the trimer. Secondary structure analysis suggests that the linker regions are composed of random coil (data not shown). This notion is supported by the H1N1 NP crystals, in which the linker regions are disordered (Ye et al., 2006), indicating that these regions are indeed flexible.

The proposed 9-mer NP structure has the following characteristics. A ring composed of 9 NPs has inner and outer diameters of 75 and 190 Å, respectively. These dimensions agree with the EM structure (Martin-Benito et al., 2001) but are slightly larger than those of the rabies and VSV NP/ RNA complexes (Albertini et al., 2006; Area et al., 2006). Alignment of the influenza virus A H5N1 NP and VSV NP sequences shows similarities in the tail loop (data not shown), yet they mediate interactions between neighboring protomers in different ways. The loops of rabies NP and VSV NP only drape along the surface of the neighboring protomer. On the other hand, the tail of the influenza virus A NP (both H5N1 and H1N1) makes extensive interaction with the adjacent protomer by inserting deeply into its body domain, and this interaction may be retained in the 9-mer model.

In the rabies and VSV NP/RNA complexes, in addition to the tail loop, the N-terminal segments also make interactions with neighboring molecules, thus helping to maintain the integrity of the ring structure (Albertini et al., 2006; Green et al., 2006). Their N-terminal 20 residues form an extended antiparallel-ribbon that reaches to

the neighboring protomer and establishes another tether. In both the H5N1 and H1N1 NP crystal structures, the 20-residue N-terminal segment is disordered, presumably pointing away from the trimer. Although this region seems to be pointing to a neighboring subunit in our model of the ring assembly (Figure 3.11A), determining whether and how an interaction is formed will require further studies.

Besides, a consequence of our model of the ring-shaped NP oligomer would be that the arginine-rich RNA binding groove is exposed on the surface, which is consistent with the finding that the bases of the RNA in the RNP are susceptible to modification (Klumpp et al., 1997).

Regions which may be involved in NP oligomerization can also be predicted from the 9-mer NP model. For instance, the helix-loop-helix motifs and the trimerization interface (aa. 421–451 and 250–272) may interact with the counterparts in the neighboring NP in a more extensive manner, when the NP molecules are more closely packed in the oligomerization state (Figure 3.11B). It is also possible that other helices in the head domain, such as helices 5 and 10, may participate in molecular contacts during NP oligomerization (Figure 3.11B).

3.2.5 Comparison of NPs from Different Negative-Sense RNA Viruses

Influenza virus, rabies virus, and VSV are all NSRVs. The sequences of their NPs differ in length (influenza: 498 residues; rabies: 398 residues; VSV: 422 residues) and show very little homology. Nevertheless, as shown in Figure 3.12, the overall conformation of influenza virus A NP is similar to that of the rabies virus (Albertini et al., 2006) and VSV (Green et al., 2006). First, the three NPs are all composed of two helical domains in a banana-shaped configuration. A long tail loop, which is

important for RNP assembly, reaches out from near the C-terminal end of the polypeptide chain and interacts with the neighboring molecule. Second, all NPs contain three loops that bridge the two domains and maintain the conformation of the protein by connecting the body and head domains. These segments in the rabies and VSV NPs are all located in the middle of the polypeptide chain but are spread throughout the sequence in the influenza virus A NP (residues 147–152, 265–278 and 450–463). Third, superimposition of the individual NP domains shows that the head domains are structurally more conserved than the body domains (Luo et al., 2007). Fourth, like VSV and rabies NPs, influenza virus A NP also contains a helix-loop-helix motif (residues 130–157), with the loop being one of the three passes between the two domains and being located at the center of the presumed RNA binding groove. These conservations in conformation among the NSRV NPs are thought to relate to their functional roles, as all of them are involved in genomic vRNA encapsidation.

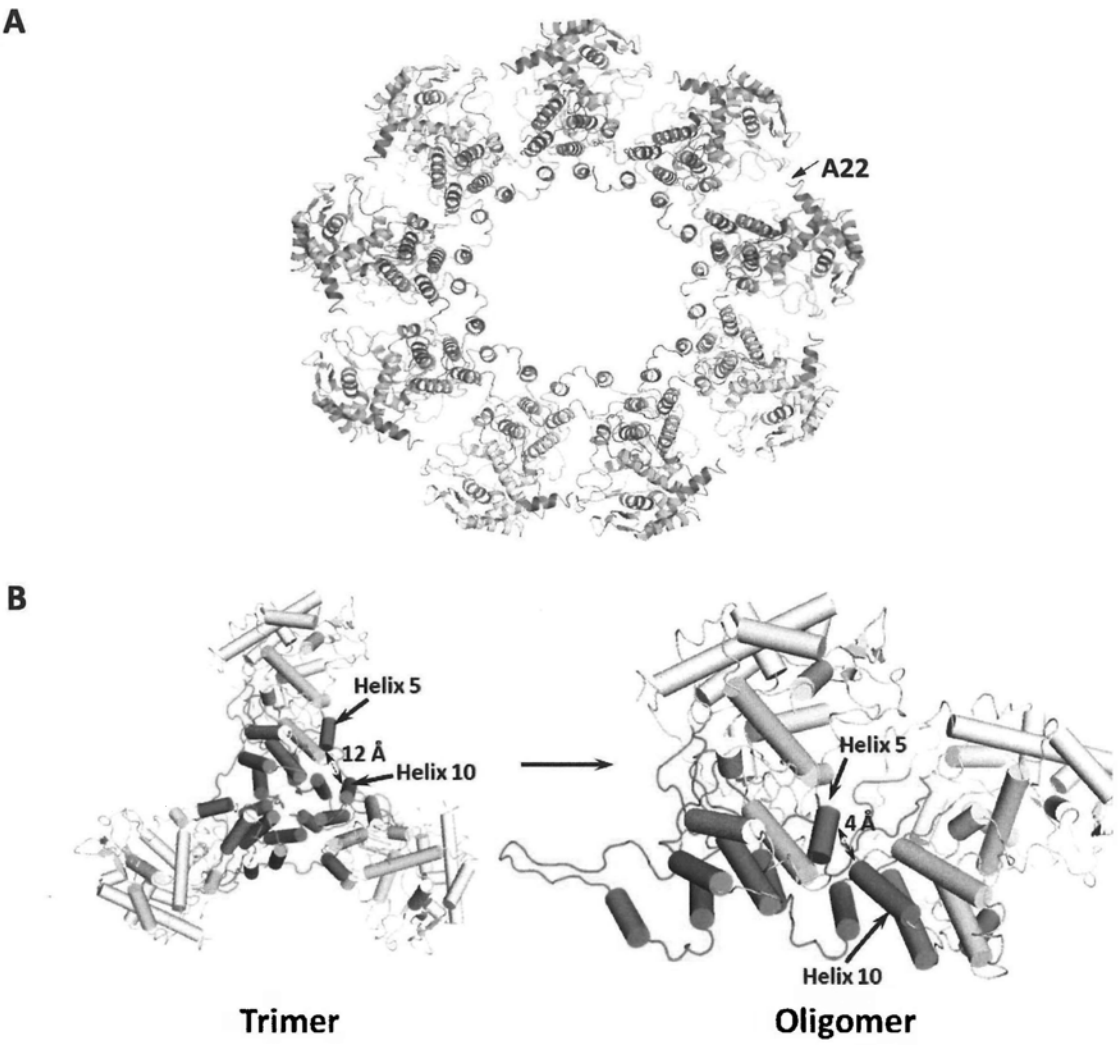


Figure 3.11 The NP 9-mer model and comparison of accessible surface area of trimeric and oligomeric NP. **(A)** Model of a ring formed by nine NP molecules based on the previously determined EM structure of a mini-RNP (Martin-Benito et al., 2001) and the flexibility of the linkers. The N-terminal region of NP is shown in red. **(B)** Helices 5 (shown in red) and 10 (shown in orange) are likely to participate in molecular contacts when trimeric NP is transformed to oligomeric NP. In the NP trimer crystal structure, helices 5 and 10 are around 12 Å apart. In the 9-mer NP model, which resembles the helical oligomeric structure in physiological conditions, the two helices are only about 4 Å apart.

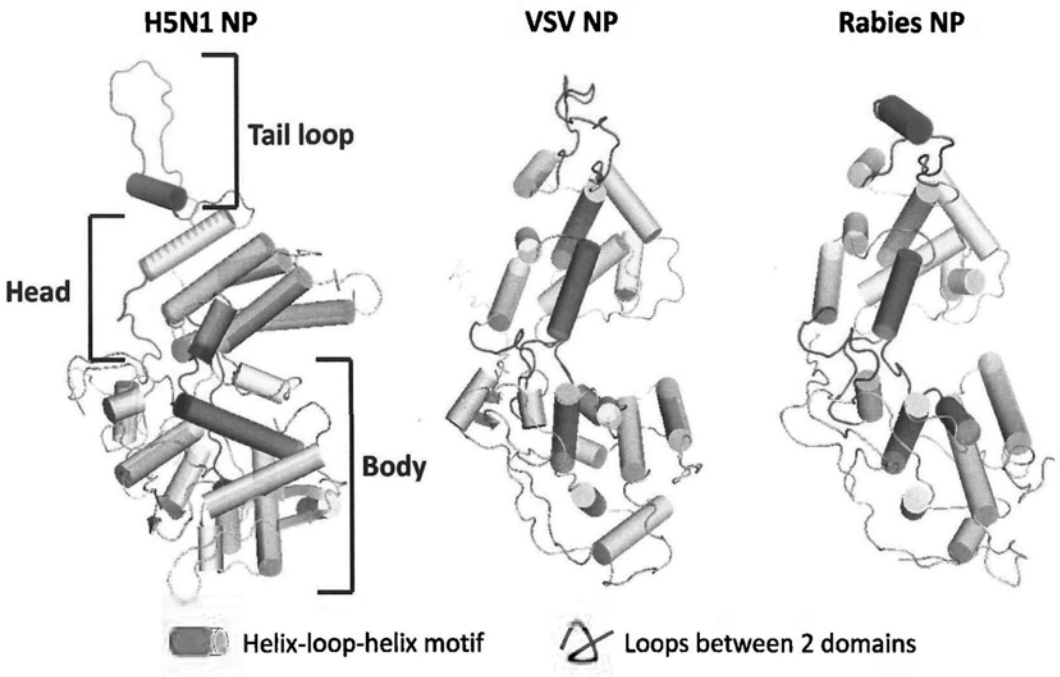


Figure 3.12 Structural comparison of the NPs from influenza virus, VSV and rabies virus. All NPs consist of head and body domains, which are connected by three segments, two loops (blue) and a helix-loop-helix motif (cyan). In all three NPs, the tail loops (red) are near the C-terminal region.

Chapter 4
Study of the Interaction between Nucleoprotein and RNA

4.1 Introduction

The role of NP in RNA binding has been investigated extensively. NP has been found to bind RNA with no sequence specificity but high affinity (Kd = 20 nM); NP also breaks the secondary structure of the RNA (Baudin et al., 1994; Yamanaka et al., 1990). Besides, chemical modification study showed that NP binds the phosphate backbone but not the nitrogenous bases of RNA (Buadin et al., 1994), in agreement with the sequence independent binding mode. UV crosslinking studies suggested that NP-RNA contact is mediated throughout the whole polypeptide (Elton et al., 1999). Chemical modification and intrinsic fluorescence microscopy have identified arginine and tryptophan instead of lysine for RNA binding (Elton et al., 1999; Medcalf et al., 1999). Before the NP structure was available, mapping of the amino acid sequences on NP that interact with RNA was by mutagenesis and biochemical techniques. However, some of the findings could not be adequately explained by the NP structure. For example, residues W120, W139, W330, F412 and R416 were found to be important for RNA encapsidation (Elton et al., 1999), but they are actually buried inside the NP structure. As revealed from the rabies NP-RNA and VSV NP-RNA structures, it is unlikely that influenza A NP complexed to RNA would have a major structural change.

NP/RNA co-crystal structures of rabies virus and VSV reveal that these NPs make use of a groove between the head and body domains for encapsidating RNA (Albertini et al., 2006; Green et al., 2006). One NP molecule of rabies virus and VSV binds 9 nucleotides of RNA. On the other hand, one influenza A NP molecule binds 24–27 nucleotides *in vivo* (Area et al., 2004; Martin-Benito et al., 2001) and tightly binds 18 or more nucleotides *in vitro* (Yamanaka et al., 1990). As shown from the crystal structure of influenza NP, it also contains such a groove as in the NPs of rabies and VSV. The aim of this study is to investigate how and where on influenza A NP binds RNA.

4.2 Results and Discussion

4.2.1 Identification of Possible RNA-Binding Regions from the Crystal Structure of NP

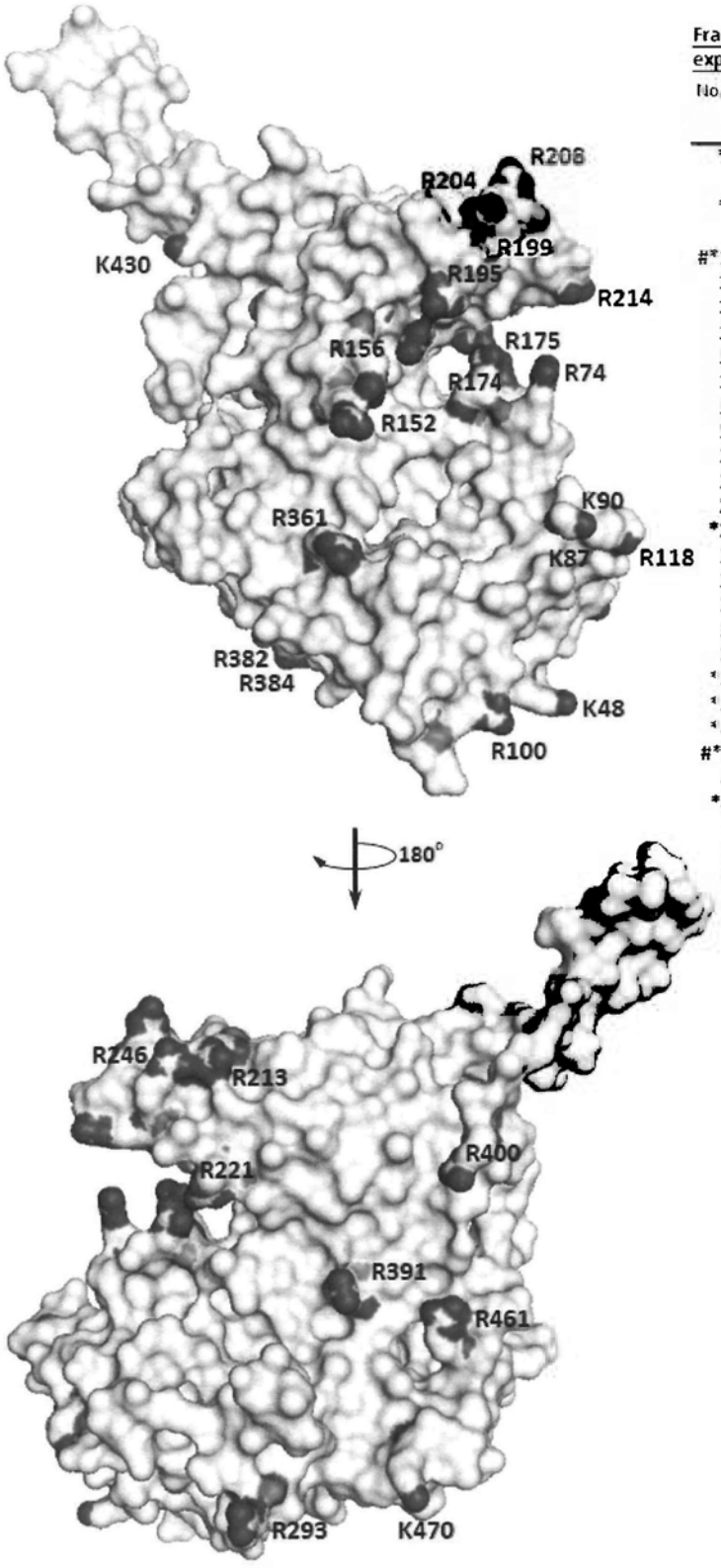
There are two unique features on influenza A virus NP which make it very different from the NP of rabies virus and VSV. First is the presence of a protruding element spanning residues 167 to 186, second is the flexible loop of residues 72 to 91, both are rich in basic residues. The protruding element situates at the groove between the head and the body domains. This protrusion is highly conserved within influenza A NP, bearing 90% of sequence conservation. The electron densities of the flexible basic loop are either weak or missing. The few traceable residues indicate that it is also situated at the groove between the head and body domains. Despite this loop is poorly resolved and structurally flexible, the loop has substantially higher sequence conservation (75%) than the overall loop structure of NP (68.1%).

NP has a total of 71 basic residues, which constitute 14.2% of the protein sequence. Out of these 71 residues, 56 of them are arginine, which is significantly higher than 5.1 % of other proteins on average (Doolittle, 1989). With reference to the structure of NP and estimated by VADAR (Willard et al., 2003), it is found that 27 basic residues (22 Arginine and 5 Lysine) are exposed in the trimeric NP (Figure 4.1). Four out of five of those lysine residues are polymorphic, which is replaced by arginine in some influenza A NP (K48R, K87R, K103R, K470R). Therefore, arginine is more conserved and is likely to be involved in RNA binding, which coincides with the previous findings (Elton et al., 1999; Medcalf et al., 1999).

Four positively-charged clusters, namely NP-G1 to NP-G4; and a flexible basic loop (aa. 74-88) were identified as potential RNA-binding regions (Figure 4.2A). SPR was then used to examine their RNA-binding affinities.

4.2.2 The NP-G1, NP-G2 and the Flexible Basic Loop are Important for RNA Binding

A series of NP mutants was generated in which multiple Arg and/or Lys residues were substituted by Ala residues, and their binding kinetics and affinities were assessed by SPR. This work identifies the important residues involved in the association of trimeric/tetrameric NP with RNA, which then results in their assembly into higher-order NP oligomers and the encapsidation of the RNA. First, different concentrations of wild-type NP were passed over a sensor chip with immobilized RNA (Figure 4.3). The measured RNA binding affinity of 23.1 ± 0.8 nM was consistent with previous studies (around 20 nM) (reviewed in Portela and Digard, 2002). A series of variants NP with different concentrations were also passed through the sensor chip surface for kinetics analysis (Figure 4.3). To assess the



Fractional accessible surface area of the exposed basic residues

No.	Residue	Fractional ASA for main chain	Fractional ASA for side chain
*48	LYS	0.85	0.93
74	ARG	0.84	0.94
*87	LYS	0.76	0.58
90	LYS	0.57	0.69
#*100	ARG	0.54	0.61
118	ARG	0.51	0.6
152	ARG	0.75	0.88
156	ARG	0.63	0.65
174	ARG	0.67	**0.78**
175	ARG	0.52	**0.56**
195	ARG	0.53	0.62
199	ARG	0.47	0.56
204	ARG	0.75	**0.87**
208	ARG	0.69	**0.69**
213	ARG	0.54	0.64
*214	ARG	0.59	0.67
221	ARG	0.52	0.62
246	ARG	0.59	0.61
293	ARG	0.76	0.78
361	ARG	0.6	**0.71**
382	ARG	0.54	**0.63**
*384	ARG	0.45	**0.52**
*391	ARG	0.68	0.79
*400	ARG	0.87	**1.02**
#*430	LYS	0.42	**0.51**
461	ARG	0.53	0.62
*470	LYS	0.77	0.92

Note:
*: polymorphic residues
#: hypervariable residues

Figure 4.1 Possible RNA binding sites of NP. 27 exposed basic residues are shown, with basic charges denoted in blue and acidic charged in red. The table shows the fractional accessible surface area of the main chain and side chain of the corresponding residues.

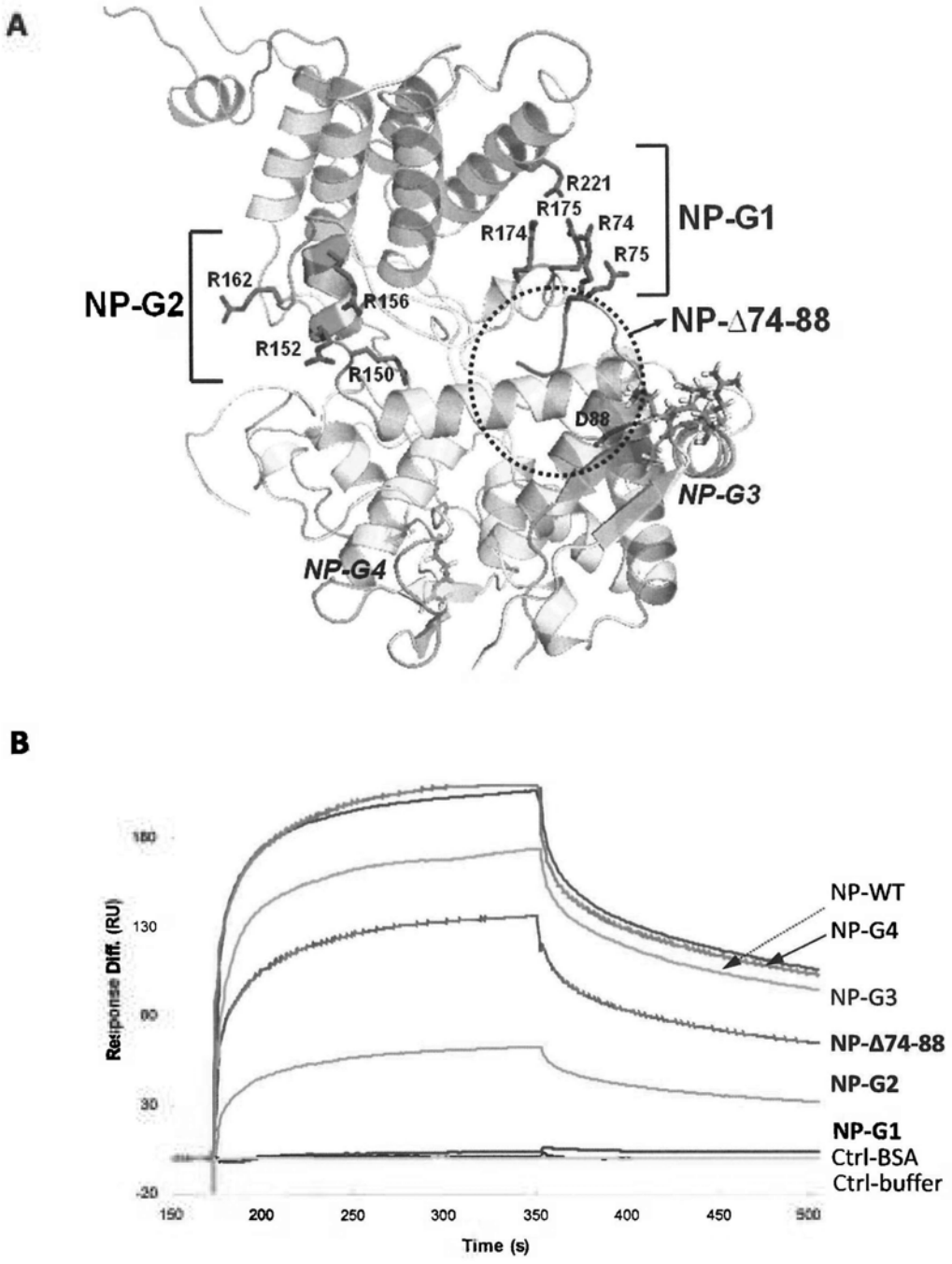


Figure 4.2 Study of the RNA-binding of NP. (A) NP regions tested for their involvement in RNA binding. Regions containing clusters of arginine and/or lysine residues are denoted as NP-G1 to NP-G4. Mutation of the arginines in NP-G1 and NP-G2 (magenta) resulted in a significant reduction in RNA binding, whereas mutations in NP-G3 and NP-G4 (green) showed little effect. **(B)** Binding of wild-type and variant NPs (700 nM) to immobilized RNA. Two controls are also shown: buffer (ctrl-buffer) and bovine serum albumin (ctrl-BSA).

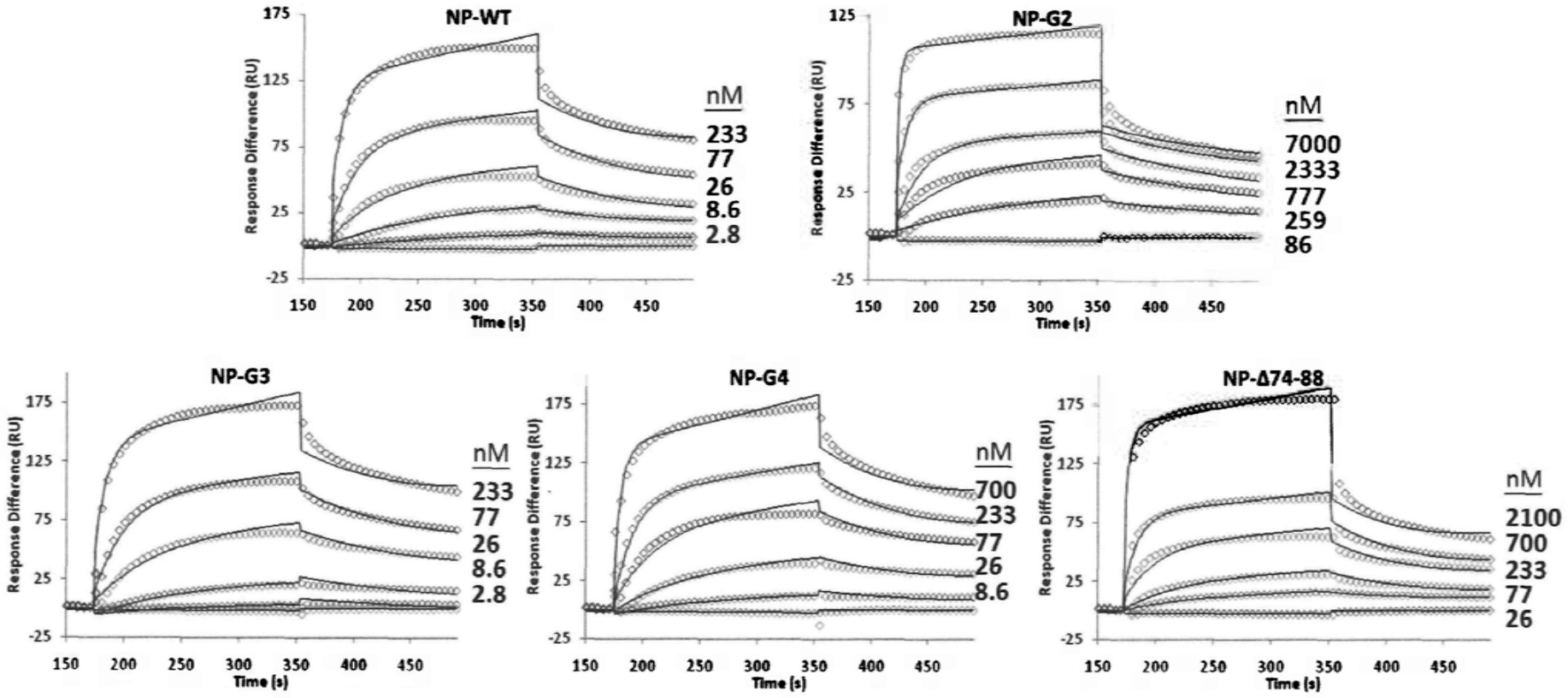


Figure 4.3 SPR graphs of the multiple-point and basic loop deletion NP variants.

Table 4.1 Kinetics and affinity constants for NP-RNA interaction

Protein	Mutation / Deletion sites	Association Rate[2] k_a (M^{-1} s^{-1})	Fold change[3]	Dissociation Rate[2] k_d (s^{-1})	Fold change[3]	Affinity[1, 2] K_D (nM)	Fold change[3]	Relative Association[4] 700 nM (RU)	70 nM (RU)	Average Binding[5] (%)
NP-WT		$4.14 \pm 0.06 \times 10^{5}$		$9.58 \pm 0.19 \times 10^{-3}$		23.1 ± 0.8		210.2	95.0	100.0
NP-G2	R150A; R152A; R156A; R162A	$4.82 \pm 0.08 \times 10^{4}$	-8.59	$5.25 \pm 0.13 \times 10^{-3}$	-0.55	108.9 ± 4.5	-4.72	62.9	19.8	25.4
NP-Δ74-88	Δ74-88	$7.84 \pm 0.16 \times 10^{4}$	-5.28	$1.16 \pm 0.03 \times 10^{-2}$	+1.21	148.0 ± 6.8	-6.41	136.1	46.6	56.9
NP-G4	R382A; R384A	$2.12 \pm 0.03 \times 10^{5}$	-1.95	$1.10 \pm 0.02 \times 10^{-2}$	+1.15	51.8 ± 1.7	-2.24	173.8	81.9	84.4
NP-G3	K90A; K91A; K113A; R117A; R121A	$3.80 \pm 0.06 \times 10^{5}$	-1.08	$9.00 \pm 0.19 \times 10^{-3}$	+0.94	23.7 ± 0.9	-1.03	206.7	96.1	99.7
NP-G1	R74A; R75A; R174A; R175A; R221A	n/d		n/d		n/d		4.3	1.5	1.8
Ctrl-BSA		n/d		n/d		n/d		1.0	0.8	0.7
Ctrl-Buffer		n/d		n/d		n/d		0.4	0.5	0.4

Note:

[1] $K_D = k_d / k_a$

[2] The errors of k_a and k_d represent the standard error while the error of K_D represents the maximum error from each experiment.

[3] Fold change is given with respect to the WT, '+' represents fold increase while '-' represents fold decrease.

[4] Value of RU was taken from t = 345 s, when NP-RNA binding has achieved equilibrium.

[5] This is calculated by averaging the percentage of RU with respect to the WT for both 700 and 70 nM protein concentrations.

contribution of the four clusters and the basic loop to RNA binding, variant NPs at concentrations of 700 and 70 nM were analyzed by SPR (Figure 4.2B and Table 4.1).

Residues R74, R75, R174, R175, and R221 in the NP-G1 region (Figure 4.2A) were found to be essential for RNA binding. Mutation of all these arginines to alanine almost completely abolished RNA binding (Table 4.1). Whereas residues R174 and R175 are part of the protrusion, R74 and R75 belong to the flexible, basic loop (residues 73-91), which contains seven basic residues (ERRNRYLEEHPSAGKDPKK). Although the electron density for this loop is not very well defined, residues 73-78 appear to extend toward the protrusion essential for RNA binding. In particular, R74 and R75 of the basic loop are in proximity to residues R174 and R175 of the protrusion. When the loop was deleted (NP-Δ74-88), the affinity of NP for RNA is 6-fold reduced (Table 4.1), demonstrating that it contributes to the ability of NP to bind vRNA, probably through an induced fit mechanism.

Another region of NP contains four arginine residues, R150, R152, R156, and R162, and is also located along the groove between the head and the body domain (NP-G2 in Figure 4.2A). It thus appeared to be another likely protein region involved in RNA binding. Changing all four arginines in the NP-G2 region to alanine caused the affinity of NP for RNA to drop by a factor of 5 (Table 4.1). Two other regions of NP also contain clusters of arginine residues: K90, K91, K113, R117, and R121 in the region denoted NP-G3, and R382 and R384 in NP-G4 (Figure 4.2A). Mutations of either cluster of arginines to alanine had little effect on RNA binding (Table 4.1).

Both regions involved in RNA binding, NP-G1 and NP-G2, contain multiple arginine residues. This feature is consistent with a previous study that employed non-specific chemical modification and found that arginine residues are more important in mediating NP-RNA contacts than lysine residues (Elton et al., 1999). The positively charged arginine residues, thought to interact with the negatively charged phosphate backbone of the RNA molecules (Elton et al., 1999), are especially enriched along the groove between the head and body domains. In our model of the ring-shaped NP assembly, this arginine-rich groove is exposed on the surface of the ring, which agrees with the RNase sensitivity of influenza virus A RNPs (Baudin et al., 1994) and the accessibility of the RNA bases to modifying agents (Klumpp et al., 1997). Moreover, a previous EM study has shown that the outside of the ring formed by influenza virus A NP interacts with RNA polymerase (Martin-Benito et al., 2001). An NP encapsidation that exposes the RNA would thus favor such interactions with the RNA polymerase. This mode of encapsidation is, however, distinctly different from the RNPs formed by the rabies virus and VSV, in which the NPs sequester the RNA at the inside of their ring-shaped assemblies (Albertini et al., 2006; Green et al., 2006). A high-resolution structure of influenza A NP-RNA oligomers will be needed to truly understand RNA encapsidation by influenza A NP.

4.2.3 Mutation of the NP-G1 and NP-G2 Regions and the Basic Loop Deletion Led to Significant Reduction in the RNP Activities

Three regions in NP (G1 [R74, R75, R174, R175, R221], G2 [R150, R152, R156, R162] and a flexible basic loop [aa. 74-88]) have now been shown to be involved in RNA binding. We then constructed multiple point arginine-to-alanine mutants targeting the G1 and G2 regions and one deletion mutant targeting the flexible basic loop, for

studying the effect of these mutations on the RNP activity *in vivo*. The mutant NP plasmids were co-transfected with plasmids expressing H5 (A/HongKong/156/97) polymerase proteins and a reporter plasmid. The amount of NP plasmid was adjusted to give similar protein expression (Figure 4.4). A plasmid expressing wild-type NP was used as the positive control, while an empty plasmid was the negative control. At 48 h post-transfection, total RNA was extracted, and the vRNA, cRNA and mRNA levels of the reporter gene were quantified by a primer extension assay, followed by polyacrylamide gel electrophoresis and autoradiography (Figure 4.5A). The various RNA levels were normalized to the internal 5S rRNA control and compared with those of the wild-type NP.

G1 and G2 mutants showed nearly undetectable activity, while the basic loop deletion mutant retained only around 10% of the RNP activity. This roughly correlates with their diminished RNA binding affinities (Table 4.1). The G1 variant was unable to bind RNA, which fits well with its undetectable RNP activity. There was a 6.4 fold decrease in the RNA binding affinity of the basic loop deletion mutant, which also correlates with its greatly reduced, but detectable, RNP activity. Surprisingly, the G2 mutant showed undetectable RNP activity yet there was only a 4.7-fold decrease in its RNA binding affinity.

We then investigated if the same phenomenon occurred with a different strain of polymerase and chose the polymerase of a popular H1N1 lab strain, A/WSN/33, to repeat the RNP reconstitution assay. The wild-type NP from H5 origin was found to be compatible with the WSN(H1) polymerase (Figure 4.5B). The RNP activities of the G1 and the Δ74-88 mutants showed the same trend as for the H5 polymerase, however the G2 mutant still retained 10% of its activity compared to wild-type

(Figure 4.5B). The RNP activity of the G2 mutant under H1 background fits well with its RNA-binding affinity (Figure 4.3, Table 4.1), but was remarkably different from that of H5 polymerase background.

Therefore, the RNA-binding affinities and the RNP activities of the NP mutants correlate with each other under the H1 background. Discrepancy of phenotypes of the NP-G2 mutant under different polymerase background was observed. This has been further studied and will be discussed in Chapter 5.

4.2.4 Mutation of the NP-G1 and NP-G2 Regions and the Basic Loop Deletion Did not Alter the NP-NP Homo-Oligomerization and NP-Polymerase Interaction

Before drawing a conclusion that the deficiency of RNA-binding leads to the defective RNP activities, we first tested if the mutants possessed intact NP-NP homo-oligomerization and NP-polymerase interaction.

To study NP homo-oligomerization *in vivo*, untagged wild-type NP and various NP mutants were co-transfected with their myc-tagged counterparts, followed by co-immunoprecipitation after 48 hours with an anti-myc antibody in the presence of RNaseA, and detection by western blotting. It was found that wild-type NP, as well as the G1, G2 and Δ74-88 mutants, were co-immunoprecipitated with their myc-tagged counterparts, however the previously described monomeric NP R416A mutant (Chan et al., 2010) did not co-immunoprecipitate (Figure 4.6A). This shows that the NP-NP homo-oligomerization ability of the NP mutants was not affected.

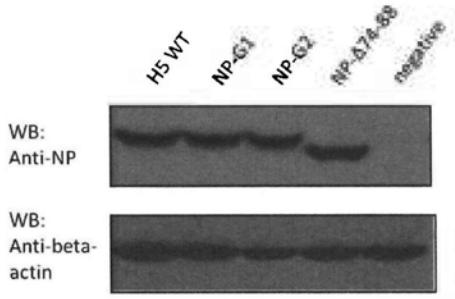


Figure 4.4 Expression of RNA-deficient NP mutants. The expression levels of the multiple point and deletion NP mutants were normalized. WB stands for Western Blotting.

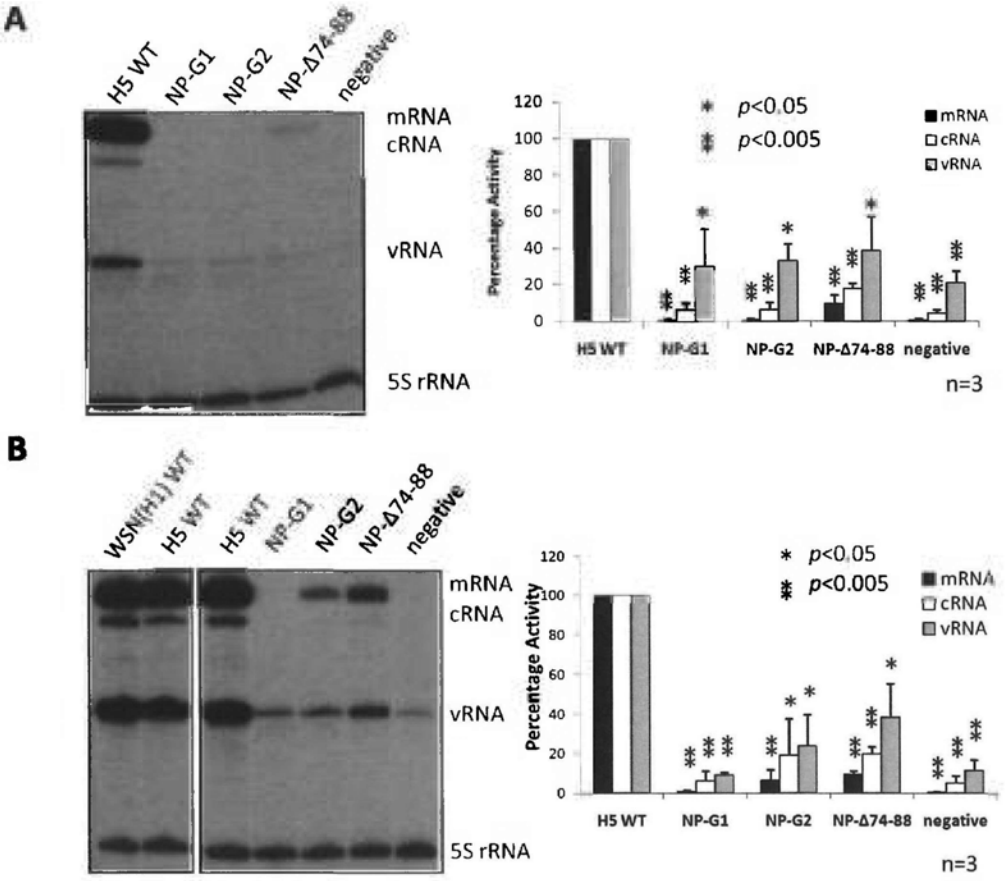


Figure 4.5 RNP reconstitution assays of NP mutants in an (A) H5 and (B) H1 polymerase background. Multiple-point and deletion NP mutants were subject to RNP reconstitution assay and their various RNA levels were quantified by primer extension. A representative result of three independent experiments is shown. RNA levels of the NP mutants were compared to those of wild-type NP, which was set to 100%. 5S rRNA was used to normalize the m-, c- and v-RNA levels. The quantitation represents the mean percentage ± standard deviations from three experiments (*, $P < 0.05$; **, $P < 0.005$).

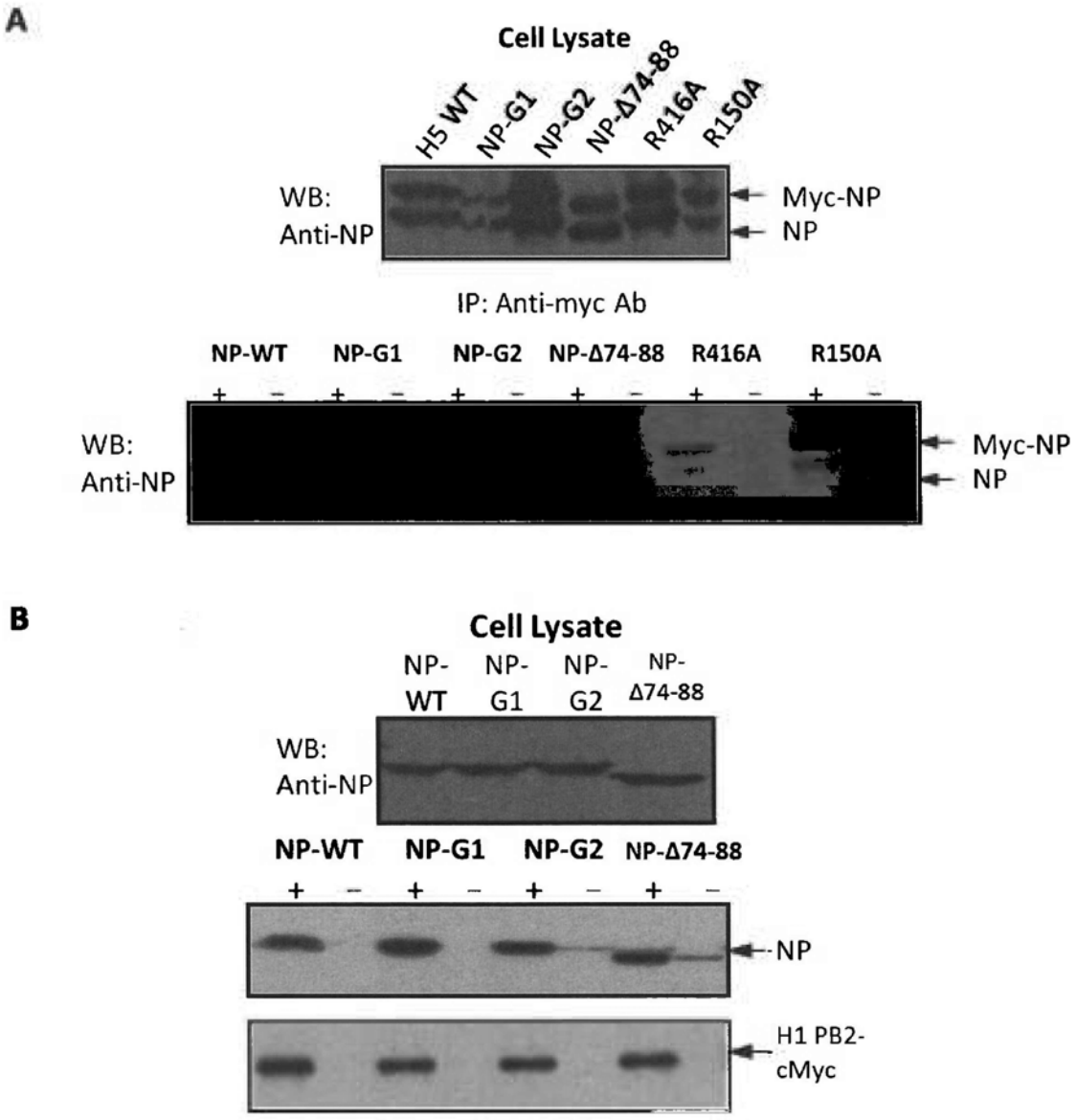


Figure 4.6 Co-immunoprecipitation experiments to study the homo-oligomerization and polymerase-binding of NP. **(A)** Myc-tagged and untagged NP mutants were expressed in 293T cells (top panel). Co-immunoprecipitation was performed with anti-Myc antibodies in the absence of RNA (bottom panel). '+' refers to the presence of the antibodies while '-' indicates their absence. R416A was used as the negative control as it is defective in NP homo-oligomerization. **(B)** Co-immunoprecipitation of NP multiple-point and deletion variants with Myc-tagged H1 polymerase. Loading controls are shown in the top panels. '+' refers to the presence of the antibodies while '-' indicates their absence.

To investigate their interaction with the viral polymerase, the NP mutants were co-expressed with the WSN(H1) myc-PB2-tagged polymerase complex and NA vRNA, co-immunoprecipitated 48 h post-transfection with anti-myc antibody and analyzed by western blotting (Figure 4.6B). It was found that the all NP mutants associated with WSN(H1) polymerases.

4.2.5 Single-Point and Double-Point Mutations of the NP-G1 and NP-G2 Regions Did not Cause Significant Reduction in RNA-Binding Affinity nor RNP Activities

Since both the G1 and G2 mutants contained multiple mutations, we set out to determine which individual residues lead to the abolishment of RNP activity. We therefore constructed nine single-point and two double-point arginine-to-alanine NP mutants for assaying their RNP activities in the H1 and H5 background. Again, their protein expression levels were normalized (Figure 4.7) and their effects on RNA levels were quantified by a primer extension assay. In the H5 polymerase background, none of the single-point or double-point mutants at the G1 cluster led to inactive RNP, suggesting that the defective phenotype of the G1 mutant was caused by multiple mutations (Figure 4.8). For the G2 cluster, a single-point mutation R150A accounted for the total loss of the RNP activity in the G2 mutant, while the three other single-point mutations, R152A, R156A and R162A, had no significant effect on RNP activity (Figure 4.9A). When the individual G2 mutations were investigated in a WSN(H1) background it was found that the R150A mutation retained 45% of its activity compared to wild-type, which is significantly different from that found in an H5 background (Figure 4.9B).

We have also individually introduced the eleven single-point and double-point mutations of the G1 and G2 clusters into a bacterial expression vector. The eleven NP mutants (R74A, R75A, [R74A,R75A], R174A, R175A, [R174A,R175A], R221A, R150A, R152A, R156A and R162A) and wild-type NP were expressed in *E. coli* and purified. We then tested the RNA binding affinities of these NP mutants using SPR (Figure 4.10, Table 4.2). No significant reduction in their RNA-binding affinities was observed.

To sum up, we have shown that the NP-G1, NP-G2 and basic loop deletion mutants had significantly reduced or undetectable RNP activities in a WSN(H1) polymerase background (Figure 4.5B), but they maintained normal NP-NP homo-oligomerization (Figure 4.6A) and NP-polymerase interaction (Figure 4.6B) abilities. We can therefore conclude that the reduced RNA binding affinities of these protein variants caused their diminished RNP activities. However, single- and double-point mutations found in the NP-G1 and NP-G2 clusters did not reduce the RNA binding affinities (Figure 4.10, Table 4.2) and the RNP activities (Figures 4.8 and 4.9B) to a large extent. The polymerase activities of some of these mutants (R74A, R150A, R152A, R156A, R174A, R175A and R221A) have also been previously investigated using luciferase reporter assay, with similar results observed (Li et al., 2009). Double-point mutants (NP-R74A,R75A and NP-R174A,R175A) did show a 2.4 and 2.8 fold decrease in their RNA-binding affinities, which were more severe than those of the respective single-point mutants. This implies the involvement of these residues in RNA binding. However, the decrease of binding affinities may be too subtle to have an impact in the RNP activities. This suggests that NP binds RNA through multiple residues, which is similar to the tail loop maintenance for NP-NP homo-oligomerization (Chan et al., 2010).

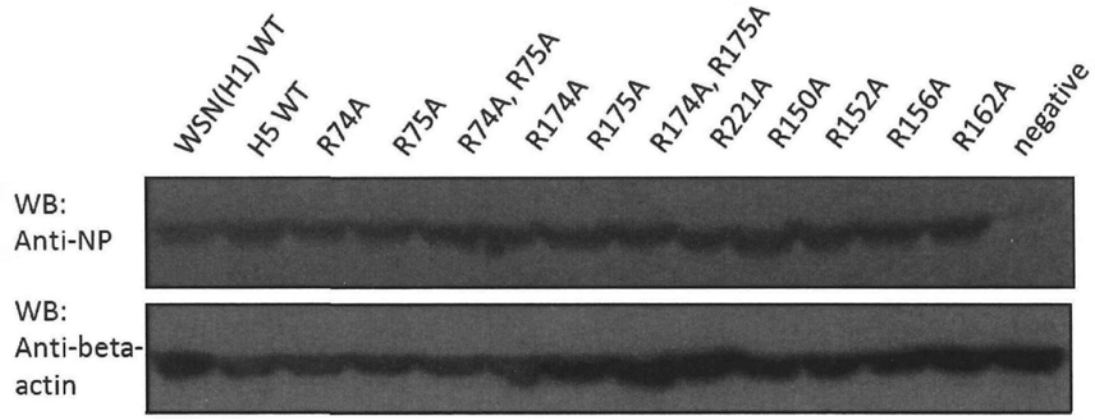


Figure 4.7 Expression of NP mutants. The expression levels of the single-point and double-point NP mutants were normalized. WB stands for Western Blotting.

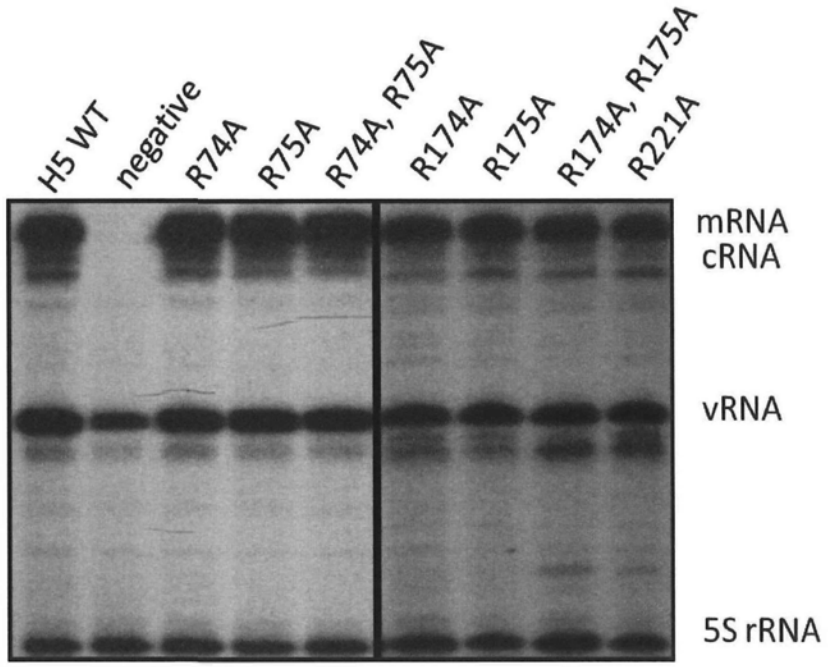


Figure 4.8 RNP reconstitution assay of single- and double- point NP mutants in the G1 cluster. Their various RNA levels were quantified by primer extension.

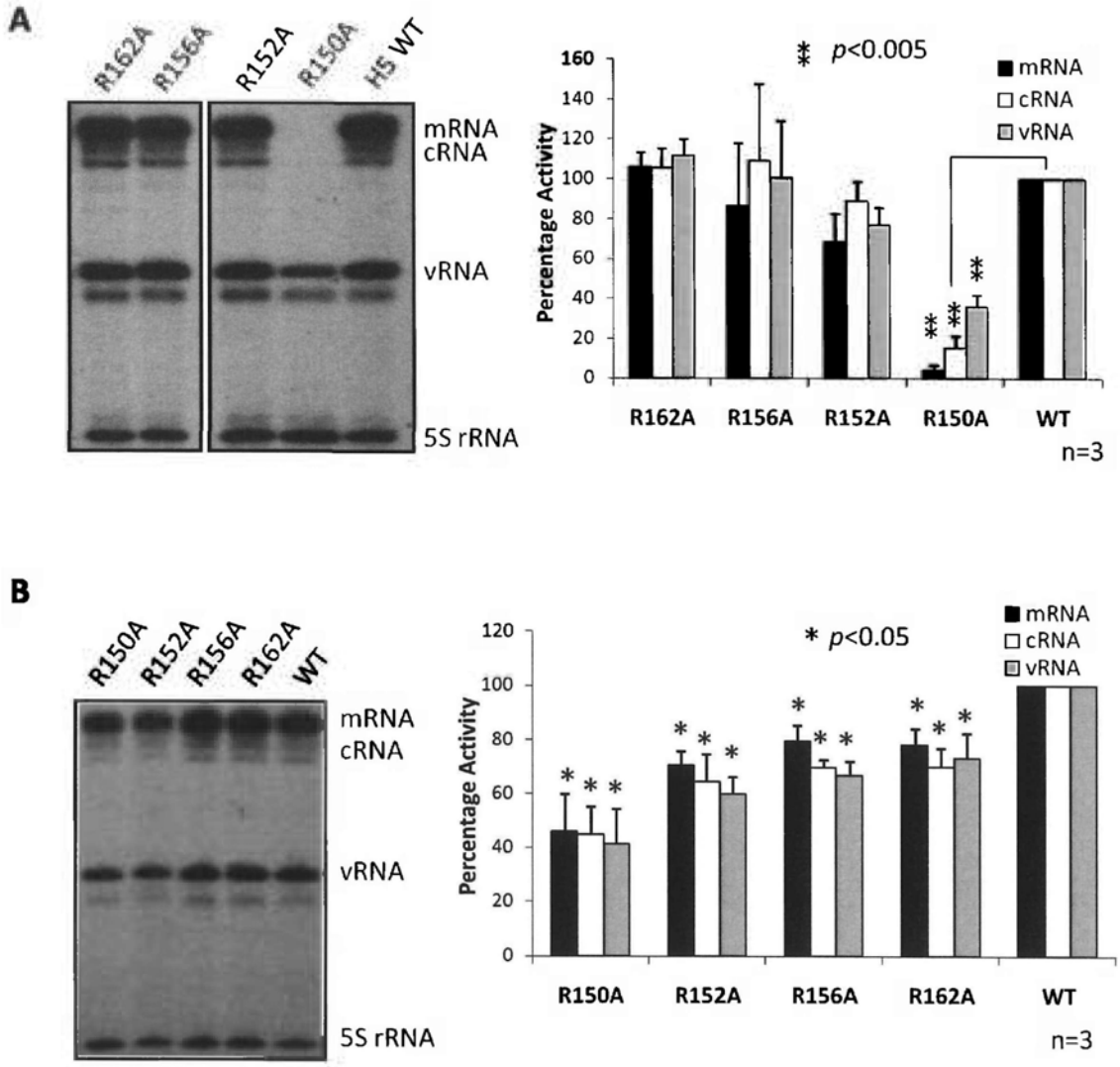


Figure 4.9 RNP reconstitution assay of single-point NP mutants in the G2 cluster in an (A) H5 and (B) H1 polymerase background. A representative result of three independent experiments is shown (*, $P < 0.05$; **, $P < 0.005$).

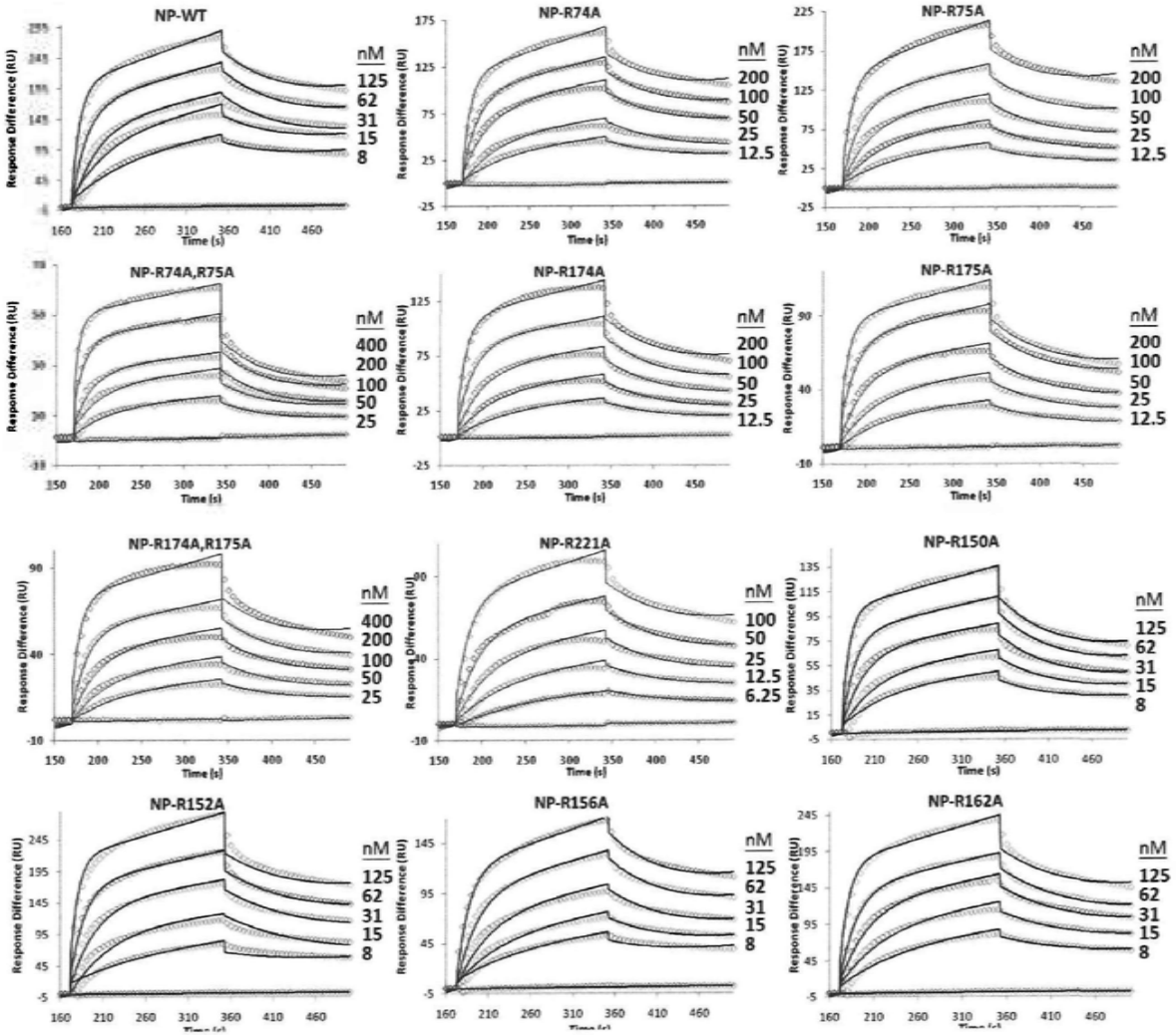


Figure 4.10 SPR graphs of the single- and double-point NP variants in the G1 and G2 clusters.

Table 4.2 Kinetics and affinity constants of NP-RNA interaction

NP	Association Rate k_a (M^{-1} s^{-1}) [2]	Dissociation Rate k_d (s^{-1}) [2]	Affinity K_D (nM) [1,2]	Fold change [3]
WT	$7.32 \pm 0.12 \times 10^{5}$	$9.96 \pm 0.22 \times 10^{-3}$	13.6 ± 0.5	-
G1 Cluster				
R74A	$4.33 \pm 0.07 \times 10^{5}$	$1.03 \pm 0.03 \times 10^{-2}$	23.8 ± 1.1	-0.8
R75A	$3.63 \pm 0.06 \times 10^{5}$	$1.08 \pm 0.03 \times 10^{-2}$	29.8 ± 1.3	-1.2
R74A,R75A	$3.13 \pm 0.05 \times 10^{5}$	$1.43 \pm 0.03 \times 10^{-2}$	45.7 ± 1.7	-2.4
R174A	$3.84 \pm 0.05 \times 10^{5}$	$1.16 \pm 0.02 \times 10^{-2}$	30.2 ± 0.9	-1.2
R175A	$4.56 \pm 0.06 \times 10^{5}$	$1.12 \pm 0.02 \times 10^{-2}$	24.6 ± 0.8	-0.8
R174A,R175A	$2.19 \pm 0.04 \times 10^{5}$	$1.12 \pm 0.03 \times 10^{-2}$	51.3 ± 2.3	-2.8
R221A	$5.23 \pm 0.07 \times 10^{5}$	$1.00 \pm 0.02 \times 10^{-2}$	19.2 ± 0.6	-0.4
G2 Cluster				
R150A	$6.60 \pm 0.09 \times 10^{5}$	$1.26 \pm 0.02 \times 10^{-2}$	19.0 ± 0.6	-0.4
R152A	$5.40 \pm 0.07 \times 10^{5}$	$1.09 \pm 0.02 \times 10^{-2}$	20.2 ± 0.6	-0.5
R156A	$4.54 \pm 0.06 \times 10^{5}$	$1.10 \pm 0.02 \times 10^{-2}$	24.2 ± 0.8	-0.8
R162A	$6.76 \pm 0.08 \times 10^{5}$	$9.59 \pm 0.16 \times 10^{-3}$	14.2 ± 0.4	0.0

Note:

[1] $K_D = k_d / k_a$

[2] The errors of k_a and k_d represent the standard error while the error of K_D represents the maximum error from each experiment.

[3] Fold change is given with respect to the WT, '+' represents fold increase while '-' represents fold decrease.

4.2.6 A Model of NP-RNA Binding

With the new information provided by the H5N1 NP structure and the SPR experiments, we can now propose a possible mechanism by which NP associates with RNA (Figure 4.11). First, the flexibility of the basic loop (residues 73–91) may allow it to sample the environment and capture RNA, explaining why deletion of the loop significantly reduces the affinity of NP for RNA. Because the loop is in close proximity to the two regions most important for RNA binding, NP-G1 and NP-G2, it could then deliver the captured RNA into the arginine-rich groove. Second, our experimental data show that the region centered around the protrusion is crucial for RNA binding and presumably is the major RNA binding site on NP. The side chains of the arginine residues in this region are pointing toward each other, suggesting that this region may clamp the RNA into the groove. Third, our data show that the NP-G2 region at the other end of the groove is also important for RNA-binding. Since 24–27 RNA nucleotides bind to an influenza virus A NP molecule (Ortega et al., 2000), as opposed to only nine RNA nucleotides in rabies and VSV NPs, the RNA is expected to make further contacts with NP in addition to binding along the arginine-rich groove.

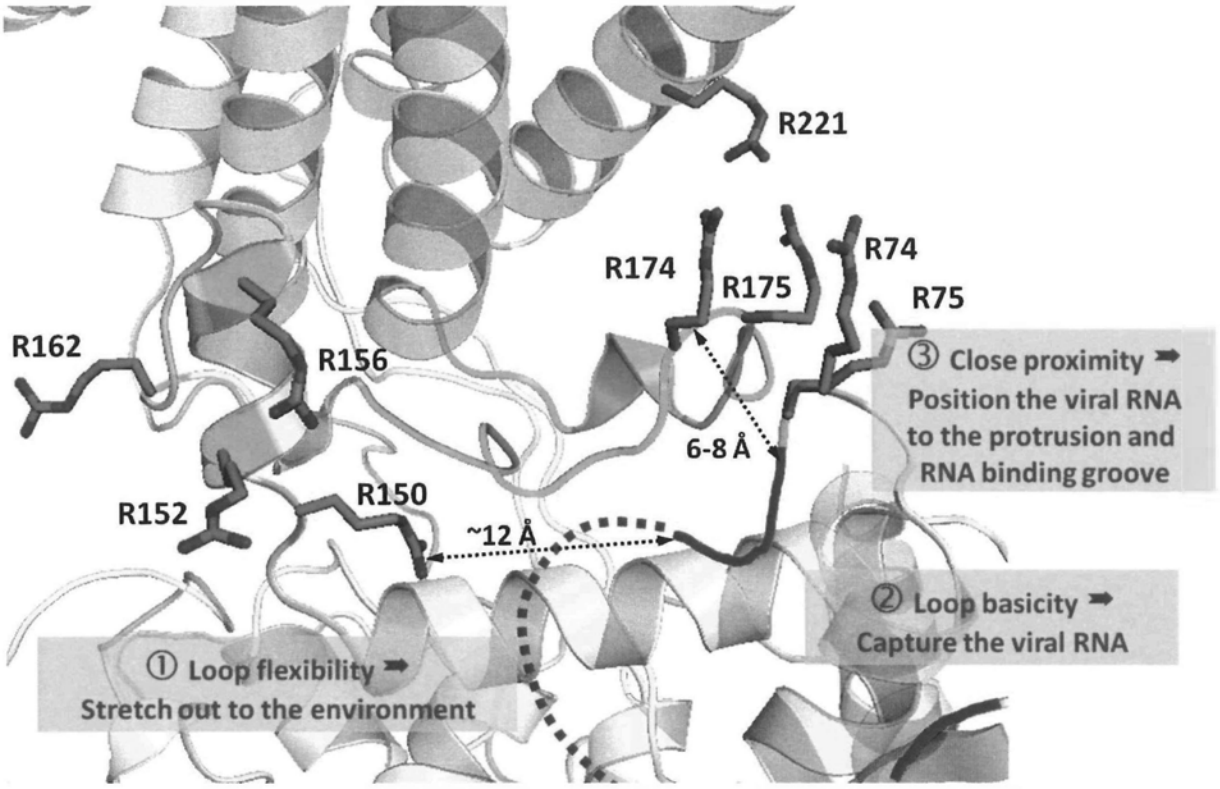


Figure 4.11 Model of NP-RNA binding. The figure shows a close-up view of the arginine-rich groove in NP. The unique protrusion (residues 167–186) found in influenza NP, but not in rabies or VSV NP, is shown in green. The distance between the side chains of R174, R175, and R221, which point toward each other, is ~ 8 Å, which is about the diameter of a single-strand RNA molecule. The flexible, basic loop (blue) is hypothesized to stretch out to the environment and capture the vRNA. The basic loop is also in close proximity to both the protrusion and the NP-G2 region and may help to position the vRNA into the basic groove.

Chapter 5
Study of the Interaction between Nucleoprotein and Polymerase Basic Protein 2

5.1 Introduction

In Chapter 4, we have identified a multiple-point NP mutant (NP-G2) that rendered the H5 polymerase totally defective but retained certain degree of activities in an H1 background (Figure 4.5). Later, we have found out the different phenotypes were due to a point mutation R150A. The R150A mutation on NP retained half of the RNP activity in an H1 background but was totally defective in an H5 background (Figure 4.9). In this chapter, I have investigated the reasons behind and shown that it was related to the NP-PB2 interaction.

The study of NP-PB2 interaction has been hindered by the poor expression of the PB2 proteins, although the interaction sites on both NP and PB2 have been roughly mapped. Three regions on NP (aa. 1-161, 255-340 and 340-465) were found to interact independently to PB2, while the C-terminal of NP (aa. 465-498) was found to inhibit NP-PB2 binding (Biswas et al., 1998). Two PB2 fragments (N-terminal aa. 1-269 and C-terminal aa. 580-683) were also identified to be responsible for NP-binding (Poole et al., 2004). However, the mapping was performed prior to the structural determination of their atomic structures. The PB2-interacting sites on NP mapped include around 75% of NP residues (Biswas et al., 1998). The mapping

was inaccurate as the NP polypeptide goes to and fro between its two domains. The two NP-binding sites covering nearly 50% of PB2 residues, were mapped by deletion mutagenesis (Poole et al., 2004). The considerably large interacting sites need to be confined.

Now, with the advancement of the protein expression screening system ESPRIT (Angelini et al., 2009), several PB2 domain structures have been determined (Guilligay et al., 2008; Tarendeau et al., 2008; Tarendeau et al., 2007). The cleared obstacle enabled us to revisit the NP-PB2 '627-domain' interaction. The re-investigation allows us to better define the interacting surfaces on the two proteins, which will be beneficial to the inhibitor design in the future.

5.2 Results and Discussion

5.2.1 NP R150A Associates with the Polymerase Complex and Forms an Active RNP in the Presence of WSN(H1) PB2, but not H5 PB2

To investigate why the defective NP R150A phenotype was strain-specific, we swapped the polymerase subunits between H5 and WSN(H1) and performed RNP reconstitution assays with either wild-type NP or the R150A mutant (Figure 5.1). All combinations of polymerases resulted in detectable RNP activity, although some were less active than others (Figure 5.1, odd numbered lanes). We speculate that this is due to differences in the compatibility of the polymerase subunits from different strains. It was found that the NP R150A mutant only resulted in RNP activity in the presence of WSN(H1) PB2 (lanes 2, 8, 10) but not H5 PB2 (lanes 4, 6, 12). PB2 is therefore the determining factor responsible for the differential phenotype of the NP R150A mutant in WSN(H1) and H5 polymerase backgrounds.

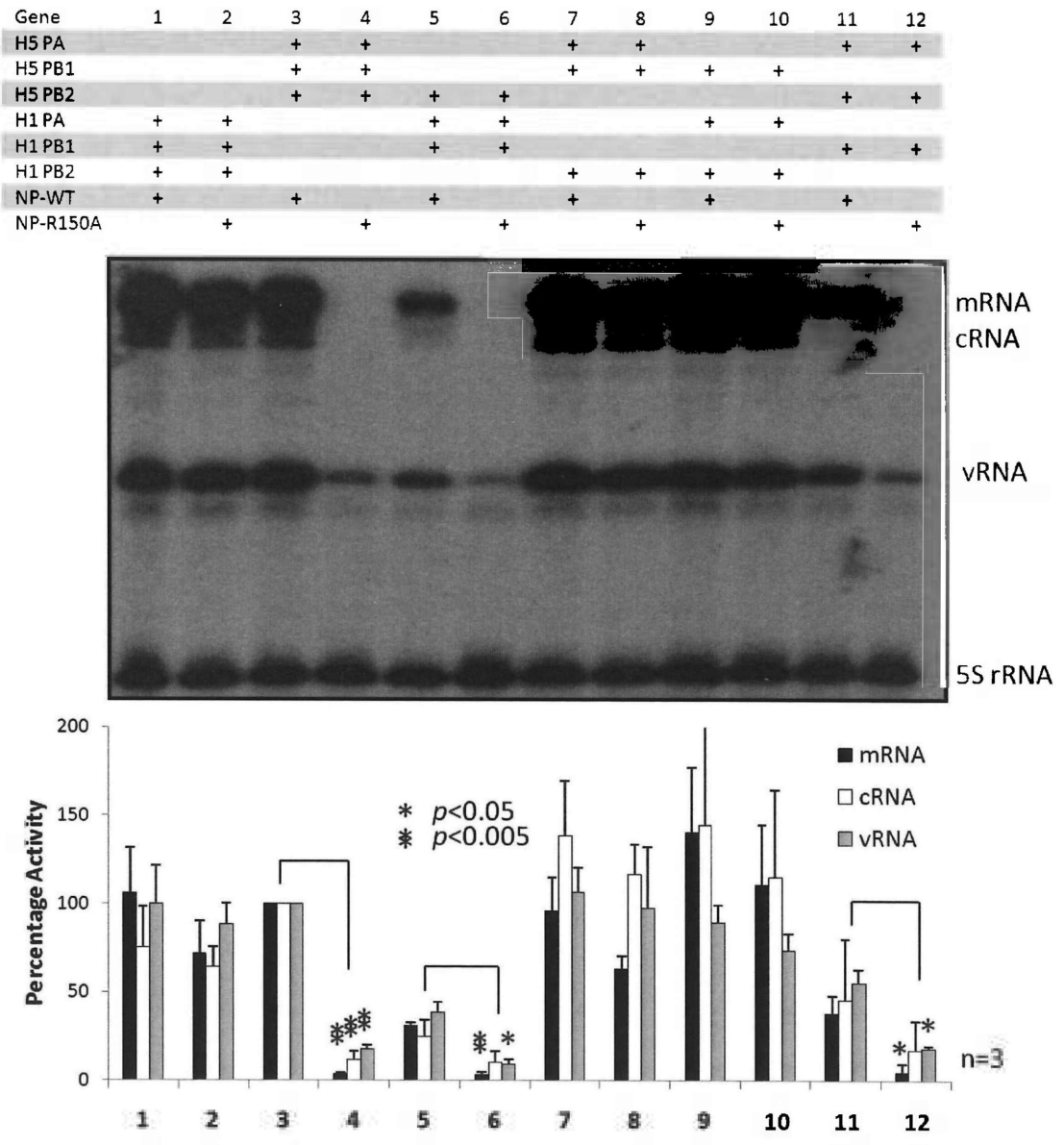


Figure 5.1 RNP reconstitution assay of gene-swapped polymerase with wild-type and R150A NP mutant. A representative result of three independent experiments is shown (*, $P < 0.05$; **, $P < 0.005$).

Why does the NP R150A mutant form an active RNP with WSN(H1) PB2 but not H5 PB2? We hypothesized that this may be related to the interaction between NP and the polymerase complex. To test this, NP wild-type or the NP R150A mutant were co-expressed with the NA vRNA, H5 PB1, WSN(H1) or H5 PB2 and myc-tagged H5 PA in 293T cells. Co-immunoprecipitation was performed at 48 h post-transfection with anti-myc antibody, followed by detection by western blotting (Figure 5.2). Wild-type NP was co-immunoprecipitated in the presence of myc-tagged polymerase complex containing either H5 or WSN(H1) PB2. However, mutation of NP R150A significantly reduced its interaction with the polymerase complex containing H5 PB2, but not WSN(H1) PB2.

Similar co-immunoprecipitation experiments were performed on the NP G1, G2 and the basic loop deletion (Δ74-88) mutants. These NP mutants were co-expressed with the H5 myc-PB2-tagged polymerase complex and NA vRNA, co-immunoprecipitated 48 h post-transfection with anti-myc antibody and analyzed by western blotting (Figure 5.3). Consistently, the NP-G2 mutant associated poorly with the H5 polymerase, in contrast to the good association with the WSN(H1) polymerase (Figure 4.6B). This supports the previous finding that the NP R150A mutant had reduced binding to the H5 PB2.

5.2.2 Substituting the WSN(H1) PB2 C-Terminus into H5 PB2 Restores the RNP Activity and NP-Polymerase Interaction of the NP R150A Mutant

Alignment of the WSN(H1) and H5 PB2 nucleotide sequences revealed an identity of 94% (results not shown). We set out to investigate which region in the WSN(H1) PB2 was responsible for interaction with NP. We therefore created a series of

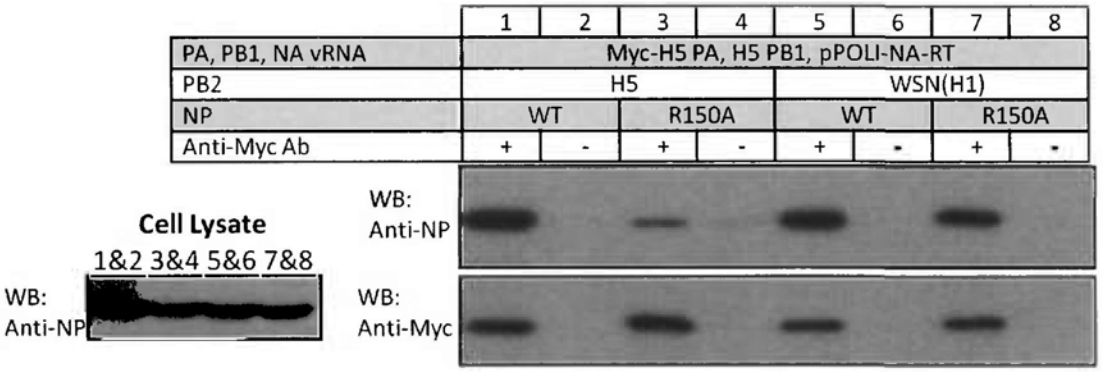


Figure 5.2 Co-immunoprecipitation of wild-type and R150A NP mutant with Myc-tagged polymerase carrying either H5 or WSN(H1) PB2. Loading controls are shown on the left.

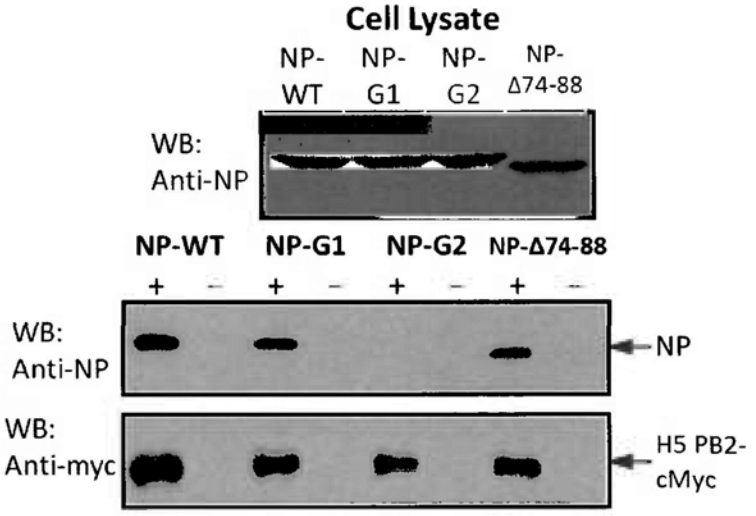


Figure 5.3 Co-immunoprecipitation of NP multiple-point and deletion variants with Myc-tagged polymerase carrying H5 PB2. Loading controls are shown in the top panels. '+' refers to the presence of the antibodies while '-' indicates their absence.

chimeric WSN(H1) and H5 PB2 proteins (Figure 5.4A). Their expression levels were first normalized (Figure 5.4B). A luciferase reporter assay (Li et al., 2009) was used to measure overall RNP activity. A plasmid encoding a vRNA-like luciferase gene was co-transfected into 293T cells with either wild-type or R150A NP, PA, PB1 and the various PB2 constructs. A plasmid encoding green fluorescent protein (GFP) was also co-transfected as a control. The RNP activity was reported as the ratio of luciferase activity to GFP levels. Transfection of all six chimeric PB2mutants with wild-type NP resulted in detectable RNP activity, indicating that they were functional (Figure 5.5A). Compared to the wild-type H5 PB2, substituting the N-terminal region of H5 PB2 with aa. 1-279 of WSN(H1) PB2 did not result in significant RNP activity with the NP R150A mutant (Figure 5.5B, H1[1-279]H5[280-759]-PB2). This indicates that the N-terminus of WSN(H1) PB2 could not recover the defective phenotype of the NP R150A mutant. On the other hand, substituting aa. 551-759 of H5 PB2 with the corresponding C-terminal region of WSN(H1) PB2 rendered the RNP active (Figure 5.5B, H5[1-550]H1[551-759]-PB2).

Co-immunoprecipitation of wild-type or R150A NP with the H5[1-550]H1[551-759]-PB2 chimeric protein was also performed. It was found that H5[1-550]H1[551-759]-PB2 could restore the NP-polymerase interaction with the NP R150A mutant (Figure 5.6). Taken together, these experiments show that the C-terminus of WSN(H1) PB2 can successfully overcome the NP R150A mutation.

5.2.3 Residues 627 and 630 in PB2 are Crucial for the NP-Polymerase Interaction

Sequence alignment of the C-terminus of PB2 of the two strains was performed to pinpoint the particular residues in the C-terminus of PB2 (aa. 551-759) which caused

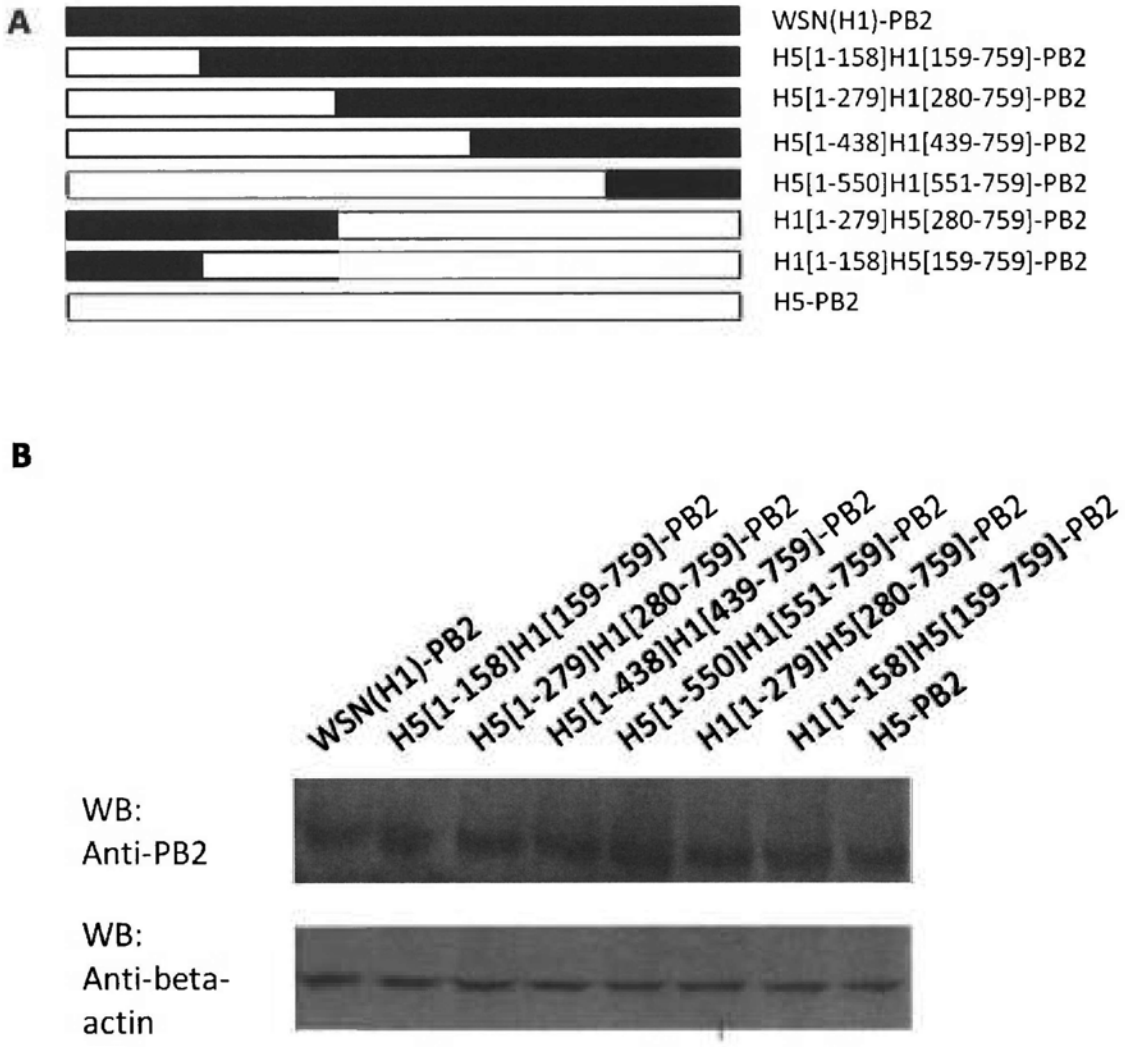


Figure 5.4 (A) Construct design and (B) expression normalization of domain-swapped PB2 mutants.

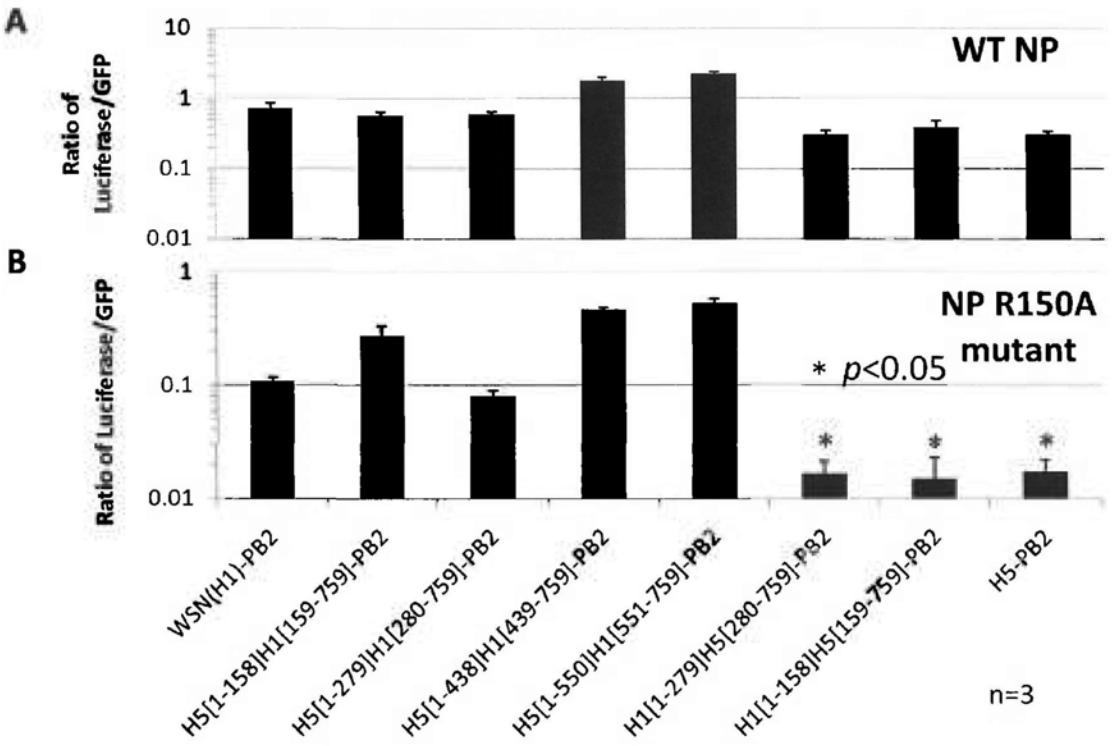


Figure 5.5 RNP reconstitution assay of domain-swapped PB2 with (A) wild-type and (B) R150A NP mutant. A plasmid encoding GFP was co-transfected for data normalization purposes. The RNP activity was reported as the ratio of luciferase activity to the GFP level. The RNP activities of the PB2 mutants were compared to that of WSN(H1)-PB2. The bar chart represents the mean ratio ± standard deviations from three independent experiments (*, $P < 0.05$).

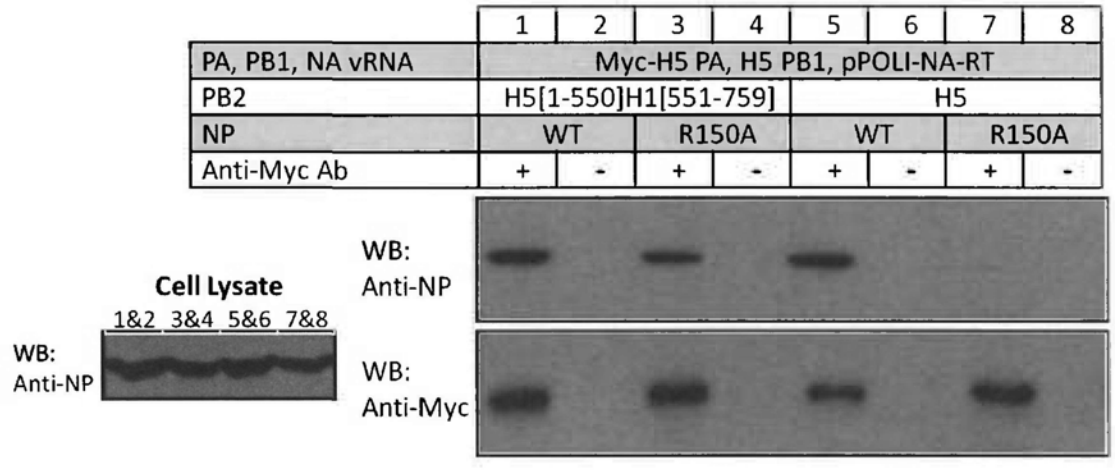


Figure 5.6 Co-immunoprecipitation of wild-type and R150A NP variant with Myc-tagged polymerase carrying either H5 or H5[1-550]H1[551-759] PB2. Loading controls are shown on the left.

the differential phenotypes in the presence of the NP R150A mutant. 13 polymorphic aa residues were found between the two strains (Figure 5.7A). We therefore introduced nine single-point and two double-point mutations into the H5 PB2, mutating the residues to those found in the WSN(H1) PB2 protein. The expression levels of these PB2 mutants were normalized (Figure 5.7B) and the activity of the mutants was measured using a luciferase reporter assay with wild-type NP (Figure 5.8A, bars 1-12) or NP R150A (Figure 5.8B, bars 1-12).

All eleven H5 PB2 mutants could form active RNPs with wild-type NP (Figure 5.8A, bars 1-12). However, only one double-point mutant ([E627K,R630G]-H5-PB2) could recover the RNP activity with NP R150A mutant (Figure 5.8B, bars 1-12). To test whether K627 or G630 of WSN(H1) PB2 recovered the RNP activity, we constructed two additional point mutants (E627K-H5-PB2 and R630G-H5-PB2). It was found that the E627K mutation in H5 PB2 gave normal RNP activity with wild-type NP and could recover the NP R150A mutation (Figures 5.8A and 5.8B, bar 14). A co-immunoprecipitation experiment also demonstrated that the E627K-H5-PB2 restored the NP-polymerase interaction of NP R150A mutant (Figure 5.9, lanes 5-8). Surprisingly, the other mutant R630G-H5-PB2, which bears E627 and G630 in H5 PB2, could not constitute significantly active RNP even with wild-type NP (Figure 5.8A, bar 15), although the protein variant was expressed in 293T cells (Figure 5.7B). We also constructed the same mutation in WSN(H1) PB2, which then contained E627 and G630 as well (K627E-H1-PB2). RNP activity with wild-type NP was greatly reduced (Figure 5.8A, bar 16). Co-immunoprecipitation experiments also showed that both mutants (R630G-H5-PB2 and K627E-H1-PB2) lost nearly all NP-polymerase interaction with wild-type NP (Figure 5.9, lanes 9-16).

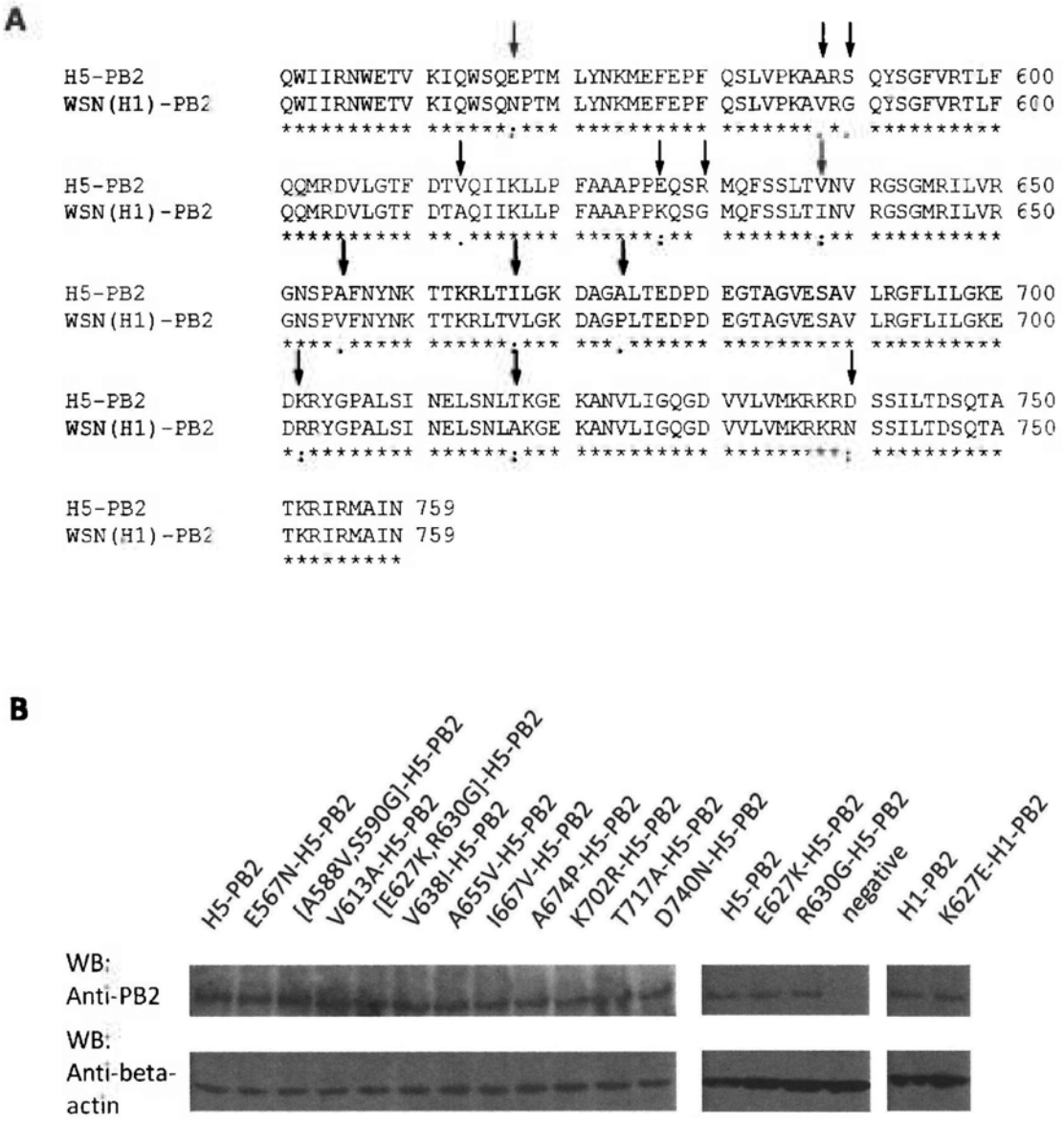


Figure 5.7 Sequence alignment and expression of mutants. (A) Sequence alignment of H5 and WSN(H1) PB2 C-terminal region (aa. 551-759). The arrows denote the differences and these H5 PB2 residues were mutated to the WSN(H1) version in the present study. (B) Expression normalization of the mutants.

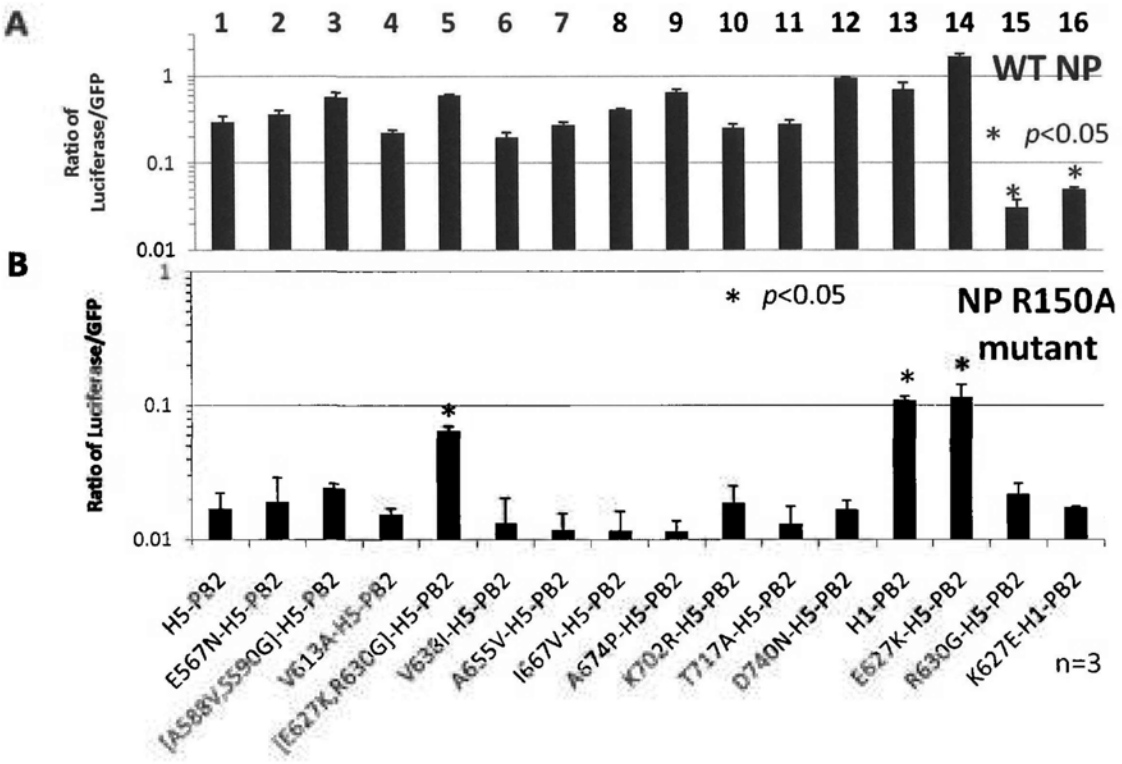


Figure 5.8 RNP reconstitution assay of PB2 point mutants with (A) wild-type and (B) R150A NP mutant. The mean RNP activities of three independent experiments of the PB2 mutants were compared to that of wild-type H5 PB2 (*, $P < 0.05$).

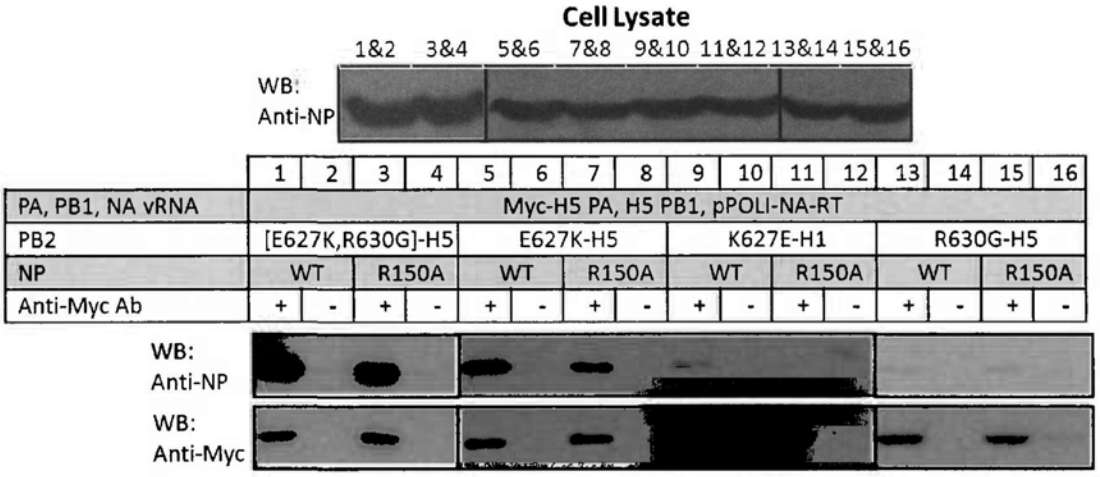


Figure 5.9 Co-immunoprecipitation of wild-type and R150A NP variant with Myc-tagged polymerase carrying four different PB2 mutants. Loading controls are shown on the top.

Since mutating either K627 or R630 resulted in an inactive RNP with wild-type NP and destroyed the NP-polymerase interaction, we conclude that both residues are important for such an interaction.

5.2.4 The '627-domain' of PB2 Directly Interacts with NP, without the Involvement of RNA

After the identification of the crucial residues in PB2 (K627 and R630) and NP (R150) required for NP-polymerase interaction, we then investigated how NP interacts with PB2, and thus the polymerase. The NP-PB2 interaction could be either direct or indirect, or even involve RNA, as both NP and PB2 have been shown to possess RNA binding activities (Ng et al., 2008; Kuzuhara et al., 2009). To understand the mode of NP-PB2 interaction, we employed *in vitro* pull-down assays and SPR with BIAcore 3000. We expressed NP and the PB2 '627-domain' (aa. 538-693) in *E. coli* and purified the protein to high homogeneity using established protocols (Ng et al., 2008; Tarendeau et al., 2008) (Figure 5.10A). The purified PB2 '627-domain' was covalently immobilized on an NHS-column. The column was then incubated with purified NP. After extensive washing, the bound protein, if any, was eluted with a high salt buffer. It was found that the PB2 '627-domain' interacts with NP (Figure 5.10B).

The purified PB2 '627-domain' was also immobilized on a CM5 chip by amine coupling. Purified NP was allowed to pass through the chip in a concentration series (Figure 5.11). The kinetic parameters were measured and the affinity of the interaction was calculated. Wild-type NP bound wild-type H5 PB2 '627-domain' through direct protein-protein interaction, with an affinity of 252 nM (Table 5.1). To determine if RNA plays a role in the interaction, a 24-nt 2'O-methylated RNA,

which confers RNase resistance, was mixed with NP in different molar ratios. The NP/RNA mixture was allowed to pass through the chip, it was shown that the NP-PB2 '627-domain' interaction decreased with increasing amounts of RNA (Figure 5.12A). This suggests that RNA is not involved in the interaction, and the presence of RNA caused an inhibitory effect.

To test if the crucial residues (PB2 K627, R630 and NP R150) for NP-polymerase binding and RNP activities are related to this direct NP-PB2 '627-domain' interaction, we cloned the [E627K,R630G] and R630G mutations separately into H5 PB2 '627-domain' for bacterial protein expression and purification (Figure 5.10B). The binding affinities of these H5 PB2 '627-domain' mutants with NP was analysed (Table 5.1). It was found that the NP R150A mutant bound 12.9 fold weaker than wild-type NP to the wild-type H5 PB2 '627-domain' (Figure 5.12B, Table 5.1). This correlates with the defective phenotype of NP R150A mutant with H5 PB2 (Figures 5.1 and 5.2). When the H5 PB2 '627-domain' [E627K,R630G] mutant was immobilized on the sensor chip, both wild-type and R150A NP variant possessed high affinities (Figure 5.12C, Table 5.1). This again correlates with the active RNP and NP-polymerase interaction of NP R150A mutant with [E627K,R630G]-H5-PB2 (Figures 5.8B and 5.9). On the other hand, neither wild-type nor the R150A NP mutant interacted with immobilized H5 PB2 R630G (Figure 5.12D). This also agrees with the lack of RNP activity and NP-polymerase interaction of R630G-H5-PB2 (Figures 5.8A, 5.8B and 5.9). Therefore, we conclude that direct protein-protein interaction of NP and PB2 '627-domain' is functionally-linked with the NP-polymerase interaction as well as the activity of the RNP complex.

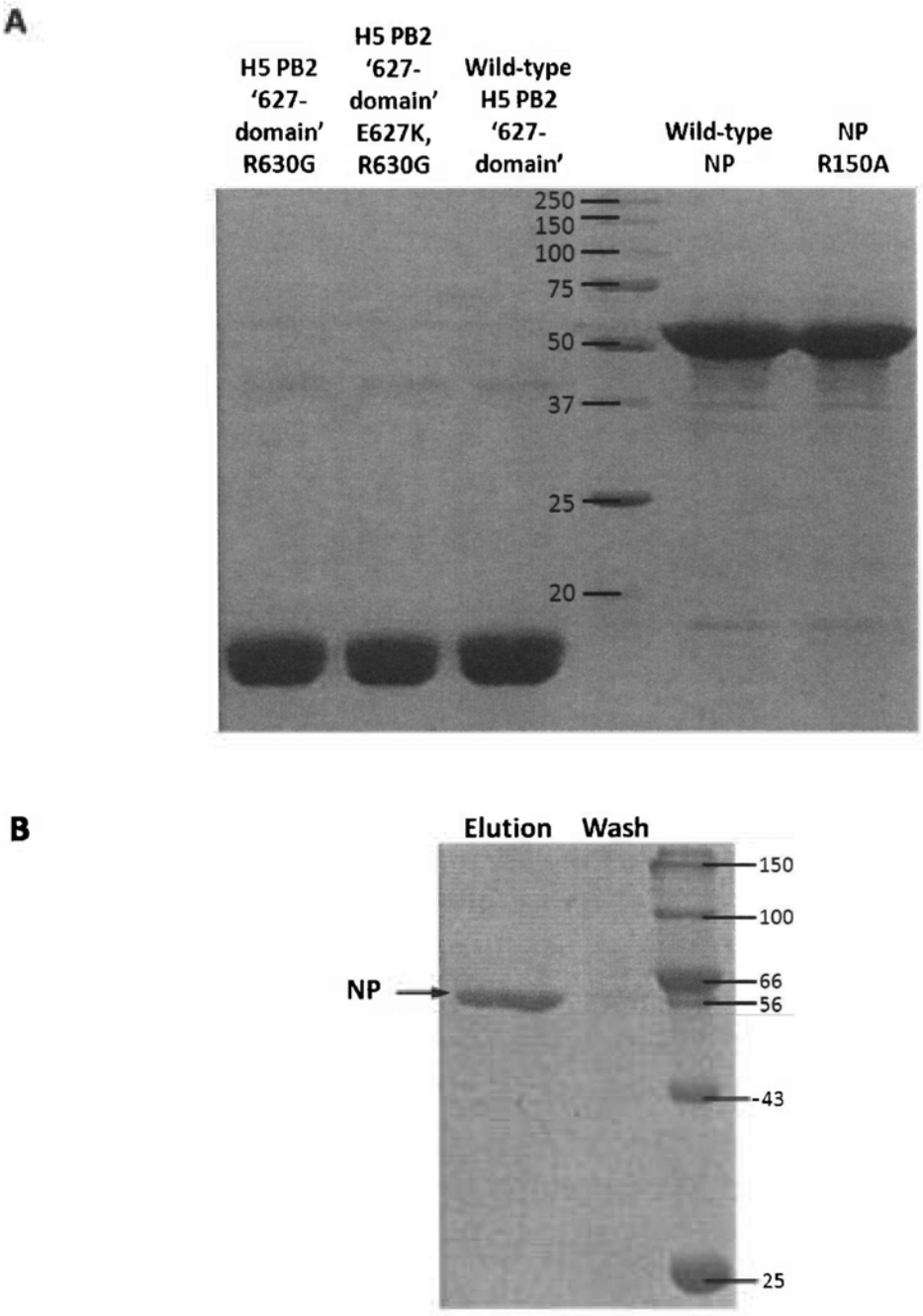


Figure 5.10 Purification of NP and PB2 '627-domain' and pull-down assay of NP-PB2 interaction. **(A)** Wild-type and variant NP and PB2 '627-domain' were purified to high homogeneity. **(B)** Purified PB2 '627-domain' was covalently immobilized onto an NHS column. Purified NP was then applied and eluted after extensive washing.

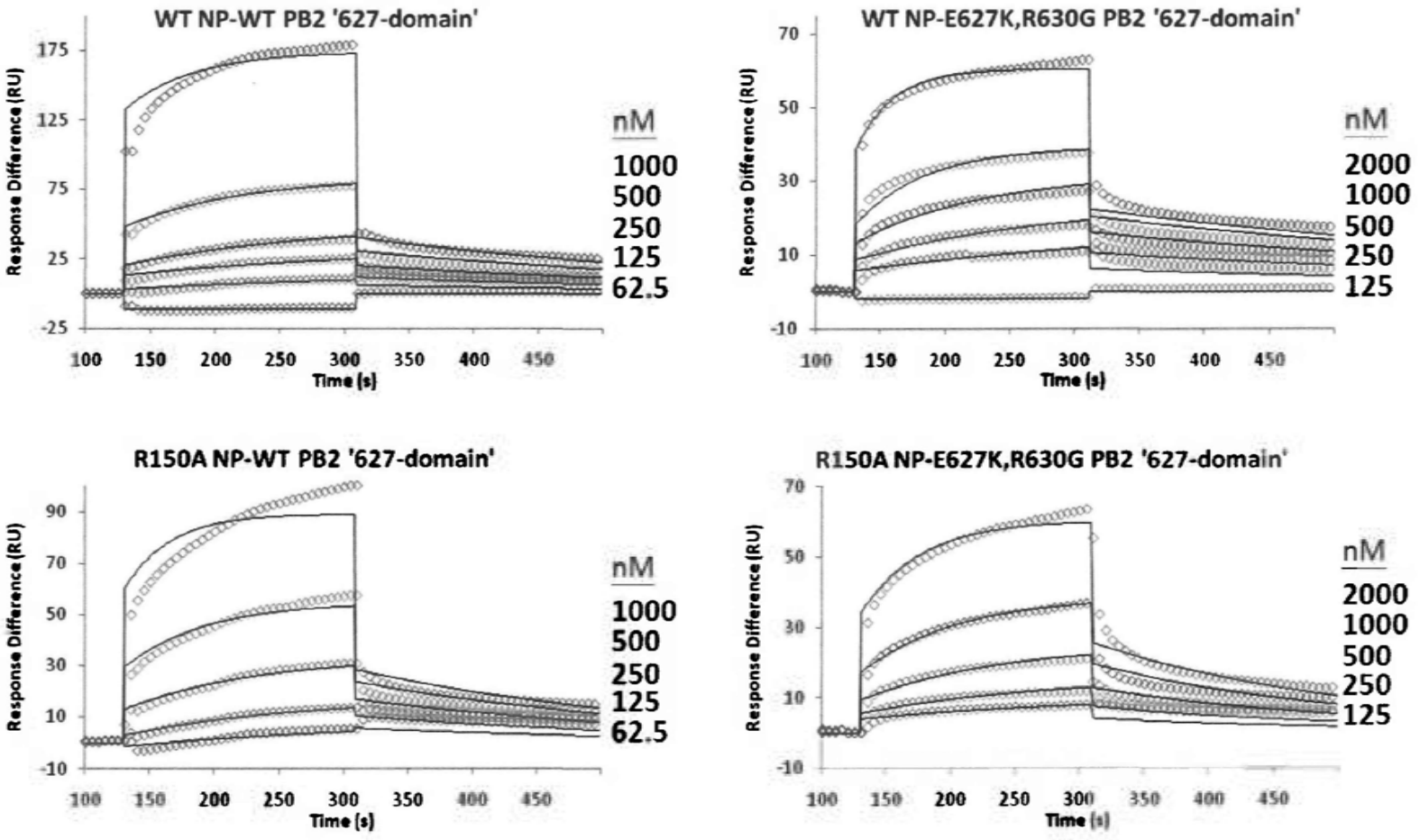


Figure 5.11 SPR graphs of NP-PB2 interaction.

Table 5.1 Kinetics and affinity constants of NP-PB2 interaction

H5 PB2	NP	Association Rate k_a (M^{-1} s^{-1}) [2]	Dissociation Rate k_d (s^{-1}) [2]	Affinity K_D (nM) [1,2]	Fold change [3]
WT	WT	$1.34 \pm 0.02 \times 10^4$	$3.37 \pm 0.08 \times 10^{-3}$	252 ± 8	-
WT	R150A	$3.26 \pm 0.12 \times 10^3$	$1.15 \pm 0.03 \times 10^{-2}$	3510 ± 39	-12.9
[E627K,R630G]	WT	$1.05 \pm 0.01 \times 10^4$	$1.76 \pm 0.06 \times 10^{-3}$	168 ± 6	+0.5
[E627K,R630G]	R150A	$6.02 \pm 0.09 \times 10^3$	$3.56 \pm 0.08 \times 10^{-3}$	591 ± 16	-1.3

Note:

[1] $K_D = k_d / k_a$

[2] The errors of k_a and k_d represent the standard error while the error of K_D represents the maximum error from each experiment.

[3] Fold change is given with respect to the WT, '+' represents fold increase while '-' represents fold decrease.

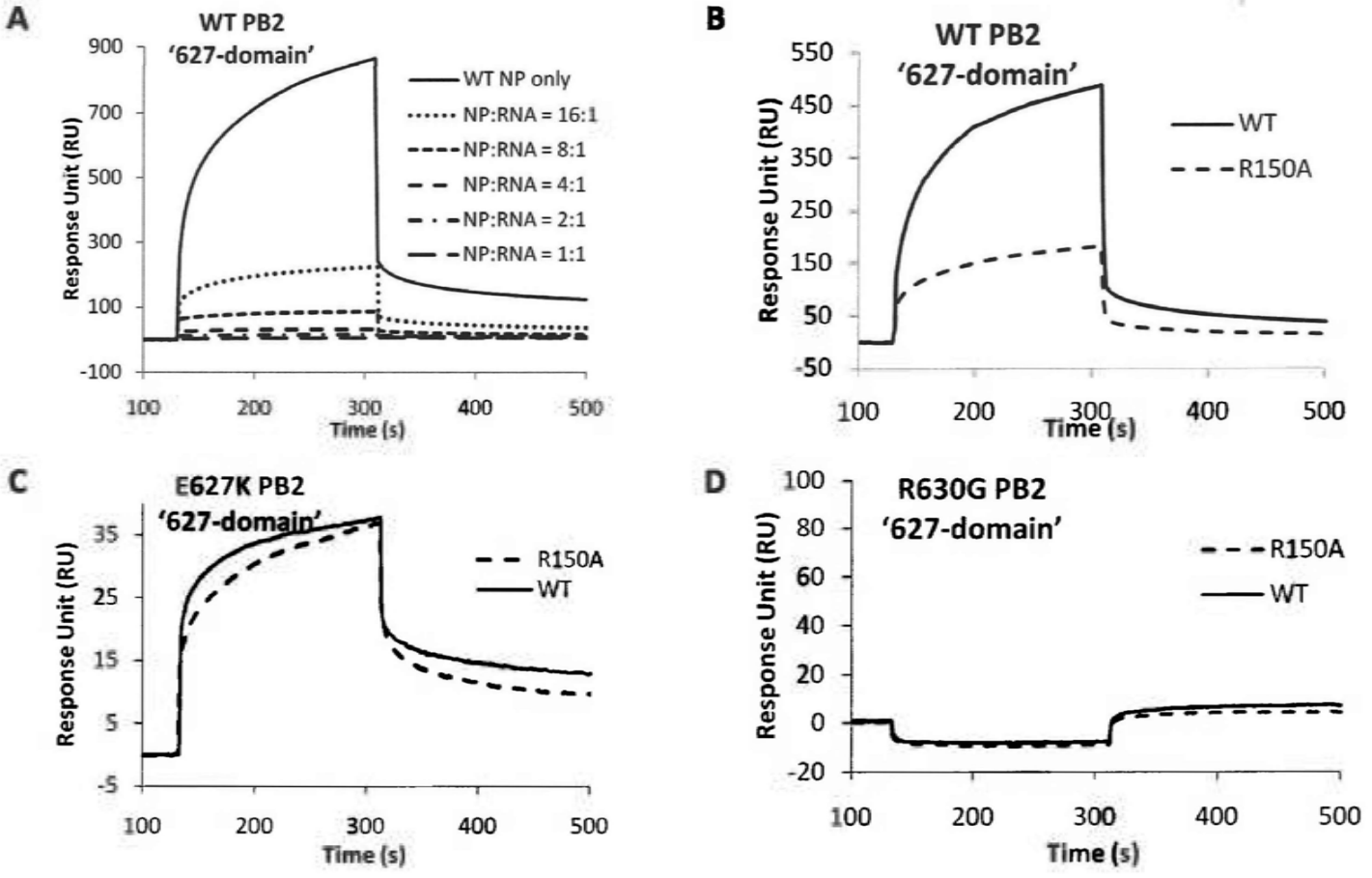


Figure 5.12 Surface plasmon resonance analysis of NP-PB2 interaction. **(A)** Wild-type PB2 '627-domain' was immobilized on a CM5 chip. Wild-type NP alone, and NP with different molar ratios of RNA were then passed through the chip. **(B, C and D)** Wild-type and R150A NP variants were analyzed with immobilized **(B)** WT, **(C)** E627K, and **(D)** R630G PB2 '627-domain' by surface plasmon resonance.

5.2.5 Dual Mechanism of NP-PB2 '627-domain' Interaction

For the NP-PB2 interaction, the important findings are summarized in Table 5.2. First, the phenotypes of the PB2 variants depend only on the residues at positions 627 and 630, but not the origin of PB2. [E627K,R630G]-H5 PB2 and WT-H1 PB2 have the same residues in positions 627 and 630 and showed similar polymerase activities and NP-polymerase interaction (rows 1-4), although they are of different origin. The same is true for R630G-H5 PB2 and K627E-H1 PB2 (rows 5-8).

Second, PB2 residues K627 and R630 can establish a functional NP-polymerase interaction independent of each other. [E627K,R630G]-H5 PB2 and R630G-H5 PB2 differ only at residue 627. However, the K627-carrying [E627K,R630G]-H5 PB2 resulted in an active polymerase and had proper NP-polymerase and NP-PB2 '627-domain' interactions with wild-type NP, while the E627-carrying R630G-H5 PB2 did not (rows 1 and 5). On the other hand, WT-H5 PB2 and R630G-H5 PB2 only differ at residue 630. Nevertheless, WT-H5 PB2, which carries R630, established a functional interaction with wild-type NP, while G630-carrying R630G-H5 PB2 did not (rows 5 and 9).

Third, PB2 K627 and R630 have different requirements for NP, with PB2 R630 requiring NP R150 for its proper function, while PB2 K627 does not. [E627K,R630G]-H5 PB2 and WT-H5 PB2, which made use of K627 and R630 to mediate the functional NP-PB2 interaction respectively, have distinct phenotypes towards the NP R150A mutant. The [E627K,R630G]-H5 PB2 mutant was capable of forming an active RNP and proper NP-polymerase interaction with the NP R150A mutant, while WT-H5 PB2 did not (rows 2 and 10).

Table 5.2 Summary of the key findings

No.	PB2			NP		Phenotype		
	Variants	aa. 627	aa. 630	Variants	aa. 150	Polymerase activity (Luciferase Assay)	NP-polymerase interaction (Co-IP)	NP-PB2 '627-domain' interaction (SPR)
1	[E627K, R630G]-H5 PB2	K	G	WT	R	Active	Yes	Yes
2				R150A	A	Active	Yes	Yes
3	WT-H1 PB2	K	G	WT	R	Active	Yes	n/a
4				R150A	A	Active	Yes	n/a
5	R630G-H5 PB2	E	G	WT	R	Inactive	No	No
6				R150A	A	Inactive	No	No
7	K627E-H1 PB2	E	G	WT	R	Inactive	Very weak	n/a
8				R150A	A	Inactive	No	n/a
9	WT-H5 PB2	E	R	WT	R	Active	Yes	Yes
10				R150A	A	Inactive	Very weak	Largely weakened

In summary, we have shown that the interaction between PB2 K627/R630 and NP is essential to the proper functioning of the RNP complex, since the disruption of the interaction renders the RNP inactive. Besides, the functional association of NP and PB2 requires one of two interactions. Both PB2 K627 and R630 can establish a functional NP-polymerase interaction independently, with the latter requiring NP R150 while the former does not. This dual interaction may confer replication advantages to influenza viruses, as either one could result in an active RNP complex.

To see if our findings correlate with naturally existing viruses, we have surveyed 10,057 PB2 sequences from the NCBI database. 61.0 % have a combination of E627 and R630, 38.2 % have K627 and R630, and the remaining 0.8 % have some minor combinations. The combination of K627 and G630, as in WSN(H1) PB2, only exists in 8 of the 10,057 PB2 sequences. The large number of E627 and R630 combinations mostly comes from the 2009 H1N1 swine influenza pandemic sequences. The natural PB2 sequences therefore support our dual mechanism of NP-PB2 interaction. It is notable that the combination of K627 and R630 is dominant over the combination of K627 and G630. When both K627 and R630 are present, the interaction may be enhanced and lead to increased RNP activity. Our luciferase data also support such a notion, as the polymerase activity of E627K-H5 PB2 with wild-type NP was significantly higher than those of WSN(H1) PB2 and H5 PB2 (p-values < 0.05 and 0.005 respectively) (Figure 5.8A, bars 1, 13, 14). This also correlated with the previous findings that the E627K mutation in an avian PB2 could strengthen the NP-polymerase interaction in human cells (Labadie et al., 2007, Rameix-Welti et al., 2009). Besides, avian viruses with E627K mutation have shown improved growth and enhanced virulence in infected cells, mice and humans (Hatta et al., 2001; Mase et al., 2006; Massin et al., 2001). Taking the previous

findings and our data together, we can provide a possible explanation for the increased virulence of avian influenza viruses carrying the E627K mutation in mammalian cells. When the PB2 protein of avian viruses carries the E627K mutation, they possess both K627 and R630. They can then utilize both residues in interacting with NP. Such a strengthened interaction boosts the polymerase activity of the virus and increases the replication advantages, therefore leading to the increased virulence.

5.2.6 Interacting Surfaces of NP-PB2 '627-domain' Interaction

Our findings also allow the identification of the NP-PB2 interacting surfaces, one at the PB2 K627 and R630 region, and the other at the NP R150 region. We have also performed molecular docking with ClusPro (Comeau et al., 2004a; Comeau et al., 2004b). One of the models may reflect the actual interaction (Figure 5.13). The negatively-charged surface of PB2 '627-domain' is shown to face the positively-charged surface of NP. When we roughly assigned the two interacting surfaces, another program, GRAMM-X (Tovchigrechko and Vakser, 2006) also gave a highly similar prediction. However, the exact molecular contacts of how the two entities interact would need further investigation.

It is likely that RNA and PB2 share the same binding site on NP. Our SPR data indicates that RNA inhibits the NP-PB2 '627-domain' interaction. This is indeed another piece of evidence suggesting the NP R150 region is a PB2-interacting surface. In the SPR, NP preferentially bound RNA rather than the PB2 '627-domain' due to their 10-fold higher affinity (Tables 4.2 and 5.1). The NP R150 region was thus masked by the RNA and was not available for PB2 binding, causing reduced NP-PB2 '627-domain' interaction with the increasing amount of RNA (Figure 5.12A).

It has previously been shown that there are two NP-binding sites on PB2, one at the N-terminal and one at the C-terminal region (Poole et al., 2004). It was unclear which region was responsible for mediating the interaction between NP and the polymerase. Our co-immunoprecipitation data now suggest that PB2 K627 and R630 at the C-terminus are crucial for NP to interact with the polymerase (Figure 5.9). The loss of RNP activities of the R630G-H5 PB2 and K627E-H1 PB2 mutants also indicated that these residues were essential for the interaction (Figure 5.8A, bars 15 and 16). There is also evidence from other groups: (1) the EM reconstruction of the mini-RNP (Coloma et al., 2009) showed that the NP R150 region is facing towards the polymerase complex. (2) The PB2 N-terminal region is unlikely to face the NP, as the EM 3D reconstruction and domain mapping of the polymerase subunit have revealed that the N-terminal PB2 is facing the environment but not NP (Area et al., 2004). (3) The N-terminal fragment (aa. 1-37) is now shown to form the PB1-PB2 complex in the crystal structure (Sugiyama et al., 2009). Taking all these together, we are convinced that NP makes a functional interaction with the polymerase by binding to PB2 K627 and/or R630. We further propose that the particular NP which interacts with PB2 in the RNP complex is not RNA-bound; instead they bind through direct protein-protein interactions, as suggested by the SPR experiment (Figures 5.12A and 5.12B).

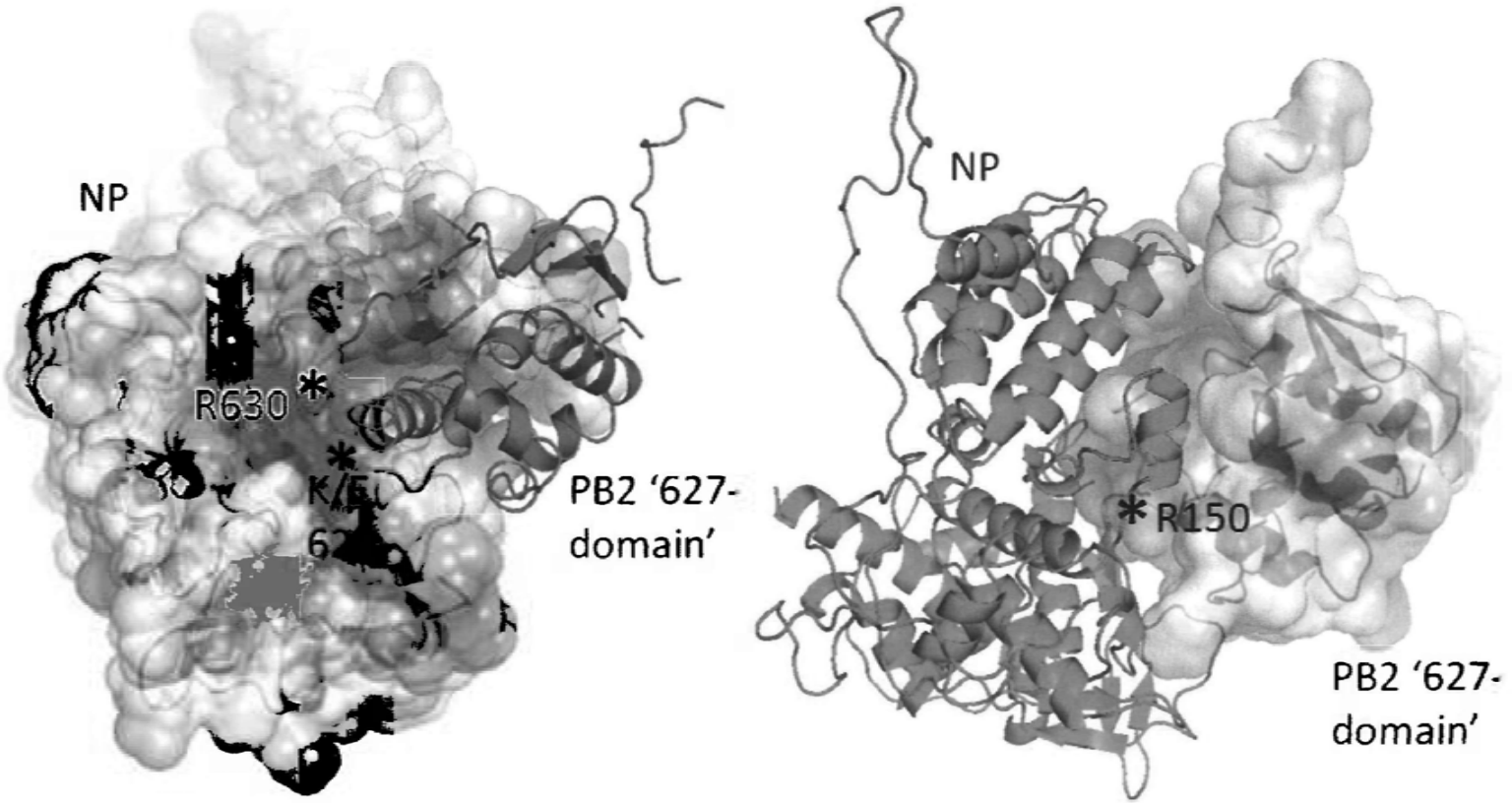


Figure 5.13 Model of NP-PB2 '627-domain' interaction predicted by ClusPro. The electrostatic surfaces of both proteins are shown. Red color indicates negative electrostatic potential while blue color indicates positive electrostatic potential. The positively-charged surface of NP (colored in green) is facing the negatively-charged surface of PB2 '627-domain' (colored in magneta). Residues K/E627 and R630 on PB2 and R150 on NP are indicated.

Chapter 6
Discussion

6.1 Concluding Remarks

In this study, we have determined the atomic structure of NP from avian origin to a resolution of 3.3 Å, which allows us to identify important structural features for NP to carry out its functions. The structural resolution urged us to revisit the molecular mechanisms of how NP interacts with other entities. We have reinvestigated the homo-oligomerization mechanism of NP (Chan et al., 2010). We have identified several crucial residues for NP-NP interaction and summarized three major forces which mediate such interaction. In this thesis, we have further our investigation into NP-RNA and NP-PB2 interactions. We have identified, by SPR and RNP reconstitution assay, three regions on NP which are important for binding RNA, which includes the NP-G1 (R74, R75, R174, R175, R221), NP-G2 (R150, R152, R156, R162), and the flexible basic loop (aa. 74-88). For the NP-PB2 interaction, by RNP reconstitution assay and co-immunoprecipitation, we have identified two residues, K627 and R630, for mediating NP-binding through independent mechanisms, with the latter requiring R150 of NP. These two residues were found to directly interact with NP, without the involvement of RNA, by SPR. Our work on NP-NP homo-oligomerization, NP-RNA binding and NP-PB2 interaction have captured evidences that influenza virus often makes use of multiple strategies for mediating its essential functions, to maximize its chance of growth and survival.

6.2 NP as a Target for Drug Design

Nucleoprotein has emerged into a new target for drug design and screening in recent years (Gong et al., 2009; Su et al., 2010). The structure of H5N1 NP should be of much value for the design of antiviral drugs to fight the next potential pandemic. It suggests that the RNA-binding groove and the PB2-interacting sites, which are exposed and highly accessible, and the tail domain, used for oligomerization of NP, would be attractive drug targets. We have shown that the RNP becomes defective when NP is incapable of binding RNA or PB2. The blocking of these sites by inhibitors should also greatly interfere with the functions of RNP. Therefore, virtual screening of small molecules which have high affinities to these sites can be performed for the identification of candidate molecules for further lead-optimization.

Recently, several compounds have been identified to bind and inhibit NP by different methods. Nucleozin, which was identified by a forward chemical genetics approach, was found to trigger NP aggregation and inhibit its nuclear accumulation (Kao et al., 2010). An artificial analog of mycalamide A, which was identified by photo-cross-linked chemical arrays, was found to bind the N-terminal aa. 1-13 of NP and inhibit the multiplication of influenza virus (Hagiwara et al., 2010). Ingavirin was found to impair the formation of mature NP oligomers and retard the nucleocytoplasmic transport (Galegov et al., 2009). However, structure-based drug screening has not been successful in identifying compounds which inhibit NP till now.

6.3 NP as a Target for Vaccine Development

Besides, a universal influenza vaccine has always been a long-term goal of

biomedical research (Gerhard et al., 2006; Livingston et al., 2006; Roose et al., 2009). The presence of $CD8^+$ T cells due to previous infection with influenza virus may reduce the severity of the pandemic (Kreijtz et al., 2008) and age-associated decline in the T-cell repertoire diversity is correlated with poor heterosubtypic protection (Yager et al., 2008). NP is a conserved internal protein and thus a good target for eliciting cell-mediated immunity. At least 14 human NP peptides have been identified as epitopes of cytotoxic T lymphocytes (CTL) (Berkhoff et al., 2005). With respect to the H5N1 NP structure, some of these epitopes, including NP380-388, NP383-391, and NP418-426, are located on random coil regions and thus susceptible to mutations, explaining why they are hypervariable (Boon et al., 2006). The accumulation of mutations in these epitopes has been associated with escape from CTL immunity (Rimmelzwaan et al., 2004). Some other epitopes, however, are in regions that are structurally and functionally important. The NP265-274 epitope, for instance, is buried inside the protein. It is part of one of the three peptide segments that span both the head and body domains and thus may be structurally important for maintaining the proper geometry of the RNA-binding groove. Another epitope, NP174-184, is located at the putative RNA-binding groove, and would thus also be functionally important for the protein (Figure 6.1). Because healthy human subjects were found to have a robust $CD4^+$ T cell response against peptides of these two conserved epitopes (Roti et al., 2008), these epitopes may be candidates for providing partial immunity to the pandemic H5N1 strain. Recently, a vaccine of modified Vaccinia virus Ankara vector encoding NP and M1, which elicits $CD8^+$ T-cell response, has completed the phase I clinical trial in health adults, and was proven to be generally safe and well tolerated (Berthoud et al., 2011).

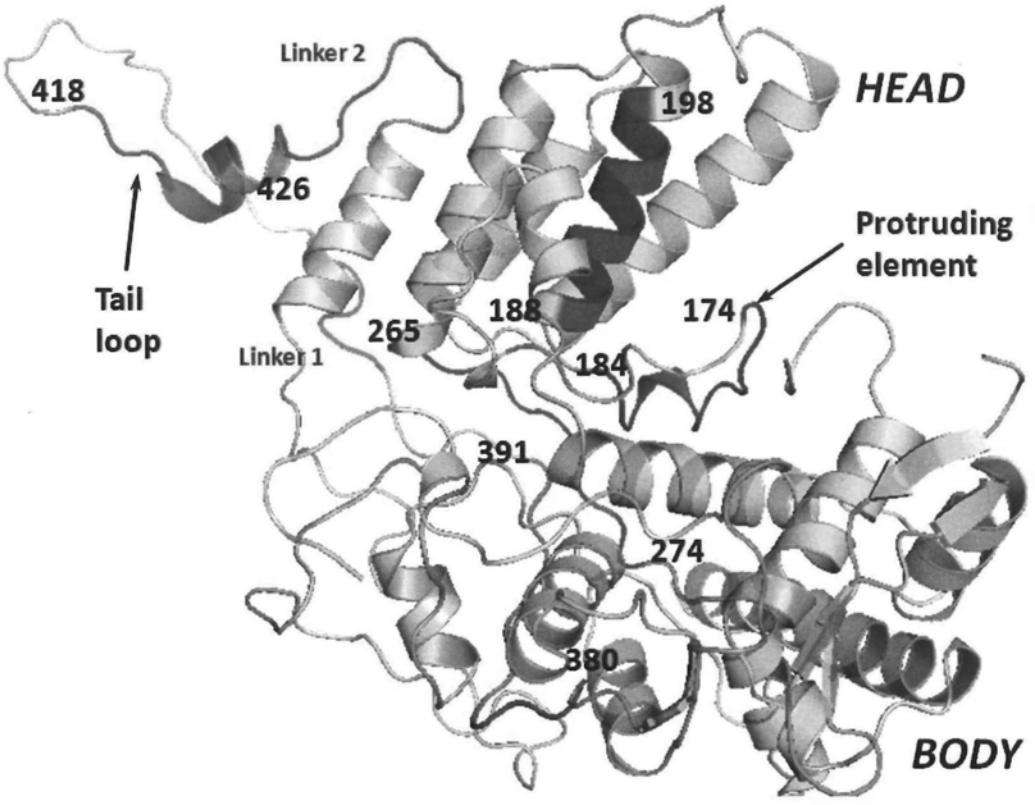


Figure 6.1 CTL epitopes on NP and their potential for vaccine development. Three epitopes, $NP_{380-388}$, $NP_{383-391}$ and $NP_{418-426}$ (colored in red), are part of a linker region or the tail loop. These three epitopes are hypervariable (Gerhard et al., 2006) and not useful for vaccine design. Three other epitopes, however, $NP_{174-184}$, $NP_{188-198}$ and $NP_{265-274}$ (colored in blue), are structurally or functionally important to NP and should therefore not be susceptible to mutations. These three epitopes should be good candidates for vaccine development.

6.4 Other Future Prospects

The study on NP-RNA binding has identified the crucial residues which mediate the process. The exact details of how NP binds RNA can only be revealed by its complex structure. Short pieces of RNA can be either soaked into pre-formed NP crystals or co-crystallized with NP proteins. The structural determination will enable the resolution of the important molecular details of the interaction and the identification of RNA-binding hot spots. This also allows the comparison of the mode of RNA-binding with the NPs of other negative-sense RNA virus, including rabies and VSV, as their complex structures with RNA have already been solved (Albertini et al., 2006; Green et al., 2006).

The identification of NP-PB2 interacting surfaces allows the further characterization of how these two entities bind with each other. Important residues other than K627 and R630 on PB2 and R150 on NP which mediate the interaction can be located by chemical shift perturbation using nuclear magnetic resonance, followed by the polymerase activity assay. As an alternative way to map the residues crucial for the interaction, chemical crosslinking coupled with mass spectrometry can also be performed. The functional significance of such an interaction can then be investigated by polymerase activity assay and virus rescue experiment. A detailed study of the interaction will definitely advance the field of influenza virus research significantly.

The atomic structure of influenza A NP could also be used as a search model for the determination of influenza B NP crystal structure. Influenza A and B NP share 37 % identities in their amino acid sequences. Still, there are significant differences between the two, which includes: (1) the much lengthened N-terminus (aa 1-67); (2) the possible RNA-binding loop (aa 124-149); (3) the dissimilar linkers (aa 452-457;

aa 485-494) and tail loop (aa 458-484); and (4) the poorly aligned C-terminus (aa 508-560). The structure of influenza B NP should also enable the structural comparison between the two. For influenza C NP, the structure of influenza A NP might or might not be useful for molecular replacement. They only share 17 % of sequence identities. Influenza C NP differs from influenza A NP by having a much lengthened C-terminus (aa. 509-565) and there are more poorly-aligned sequences in the middle of the protein. These may lead to the difficulties in molecular replacement. However, it is also possible that influenza A and C NP actually have similar folds albeit their dissimilar amino acid sequences.

Copyright Notice

Literature Cited

1. Albertini, A.A., Wernimont, A.K., Muziol, T., Ravelli, R.B., Clapier, C.R., Schoehn, G., Weissenhorn, W. and Ruigrok, R.W. (2006) Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science*, **313**, 360-363.
2. Albo, C., Valencia, A. and Portela, A. (1995) Identification of an RNA binding region within the N-terminal third of the influenza A virus nucleoprotein. *J Virol*, **69**, 3799-3806.
3. Angelini, A., Tosi, T., Mas, P., Acajjaoui, S., Zanotti, G., Terradot, L. and Hart, D.J. (2009) Expression of Helicobacter pylori CagA domains by library-based construct screening. *Febs J*, **276**, 816-824.
4. Area, E., Martin-Benito, J., Gastaminza, P., Torreira, E., Valpuesta, J.M., Carrascosa, J.L. and Ortin, J. (2004) 3D structure of the influenza virus polymerase complex: localization of subunit domains. *Proc Natl Acad Sci U S A*, **101**, 308-313.
5. Avalos, R.T., Yu, Z. and Nayak, D.P. (1997) Association of influenza virus NP and M1 proteins with cellular cytoskeletal elements in influenza virus-infected cells. *J Virol*, **71**, 2947-2958.
6. Bancroft, C.T. and Parslow, T.G. (2002) Evidence for segment-nonspecific packaging of the influenza a virus genome. *J Virol*, **76**, 7133-7139.
7. Baudin, F., Bach, C., Cusack, S. and Ruigrok, R.W. (1994) Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *Embo J*, **13**, 3158-3165.
8. Beaton, A.R. and Krug, R.M. (1981) Selected host cell capped RNA fragments prime influenza viral RNA transcription in vivo. *Nucleic Acids Res*, **9**, 4423-4436.
9. Beaton, A.R. and Krug, R.M. (1986) Transcription antitermination during influenza viral template RNA synthesis requires the nucleocapsid protein and the absence of a 5' capped end. *Proc Natl Acad Sci U S A*, **83**, 6282-6286.
10. Berkhoff, E.G., de Wit, E., Geelhoed-Mieras, M.M., Boon, A.C., Symons, J., Fouchier, R.A., Osterhaus, A.D. and Rimmelzwaan, G.F. (2005) Functional constraints of influenza A virus epitopes limit escape from cytotoxic T lymphocytes. *J Virol*, **79**, 11239-11246.
11. Berthoud, T.K., Hamill, M., Lillie, P.J., Hwenda, L., Collins, K.A., Ewer, K.J.,

Milicic, A., Poyntz, H.C., Lambe, T., Fletcher, H.A., Hill, A.V. and Gilbert, S.C. (2011) Potent CD8+ T-cell immunogenicity in humans of a novel heterosubtypic influenza A vaccine, MVA-NP+M1. *Clin Infect Dis*, **52**, 1-7.

12. Bishop, D.H., Obijeski, J.F. and Simpson, R.W. (1971) Transcription of the influenza ribonucleic acid genome by a virion polymerase. I. Optimal conditions for in vitro activity of the ribonucleic acid-dependent ribonucleic acid polymerase. *J Virol*, **8**, 66-73.
13. Biswas, S.K. and Nayak, D.P. (1994) Mutational analysis of the conserved motifs of influenza A virus polymerase basic protein 1. *J Virol*, **68**, 1819-1826.
14. Biswas, S.K., Boutz, P.L. and Nayak, D.P. (1998) Influenza virus nucleoprotein interacts with influenza virus polymerase proteins. *J Virol*, **72**, 5493-5501.
15. Blaas, D., Patzelt, E. and Kuechler, E. (1982) Identification of the cap binding protein of influenza virus. *Nucleic Acids Res*, **10**, 4803-4812.
16. Boivin, S., Cusack, S., Ruigrok, R.W. and Hart, D.J. (2010) Influenza A virus polymerase: structural insights into replication and host adaptation mechanisms. *J Biol Chem*, **285**, 28411-28417.
17. Boon, A.C., de Mutsert, G., Fouchier, R.A., Osterhaus, A.D. and Rimmelzwaan, G.F. (2006) The hypervariable immunodominant NP418-426 epitope from the influenza A virus nucleoprotein is recognized by cytotoxic T lymphocytes with high functional avidity. *J Virol*, **80**, 6024-6032.
18. Bourmakina, S.V. and Garcia-Sastre, A. (2003) Reverse genetics studies on the filamentous morphology of influenza A virus. *J Gen Virol*, **84**, 517-527.
19. Bourmakina, S.V. and Garcia-Sastre, A. (2005) The morphology and composition of influenza A virus particles are not affected by low levels of M1 and M2 proteins in infected cells. *J Virol*, **79**, 7926-7932.
20. Braam, J., Ulmanen, I. and Krug, R.M. (1983) Molecular model of a eucaryotic transcription complex: functions and movements of influenza P proteins during capped RNA-primed transcription. *Cell*, **34**, 609-618.
21. Brownlee, G.G. and Sharps, J.L. (2002) The RNA polymerase of influenza a virus is stabilized by interaction with its viral RNA promoter. *J Virol*, **76**, 7103-7113.
22. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*, **54**, 905-921.
23. Bui, M., Whittaker, G. and Helenius, A. (1996) Effect of M1 protein and low pH on nuclear transport of influenza virus ribonucleoproteins. *J Virol*, **70**,

8391-8401.

24. Chan, W.H., Ng, A.K., Robb, N.C., Lam, M.K., Chan, P.K., Au, S.W., Wang, J.H., Fodor, E. and Shaw, P.C. (2010) Functional analysis of the influenza virus H5N1 nucleoprotein tail loop reveals amino acids that are crucial for oligomerization and ribonucleoprotein activities. *J Virol*, **84**, 7337-7345.
25. Chase, G., Deng, T., Fodor, E., Leung, B.W., Mayer, D., Schwemmle, M. and Brownlee, G. (2008) Hsp90 inhibitors reduce influenza virus replication in cell culture. *Virology*, **377**, 431-439.
26. Chen, W., Calvo, P.A., Malide, D., Gibbs, J., Schubert, U., Bacik, I., Basta, S., O'Neill, R., Schickli, J., Palese, P., Henklein, P., Bennink, J.R. and Yewdell, J.W. (2001) A novel influenza A virus mitochondrial protein that induces cell death. *Nat Med*, **7**, 1306-1312.
27. Cianci, C., Tiley, L. and Krystal, M. (1995) Differential activation of the influenza virus polymerase via template RNA binding. *J Virol*, **69**, 3995-3999.
28. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, **50**, 760-763.
29. Coloma, R., Valpuesta, J.M., Arranz, R., Carrascosa, J.L., Ortin, J. and Martin-Benito, J. (2009) The structure of a biologically active influenza virus ribonucleoprotein complex. *PLoS Pathog*, **5**, e1000491.
30. Comeau, S.R., Gatchell, D.W., Vajda, S. and Camacho, C.J. (2004) ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res*, **32**, W96-99.
31. Comeau, S.R., Gatchell, D.W., Vajda, S. and Camacho, C.J. (2004) ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*, **20**, 45-50.
32. Compans, R.W., Content, J. and Duesberg, P.H. (1972) Structure of the ribonucleoprotein of influenza virus. *J Virol*, **10**, 795-800.
33. Connor, R.J., Kawaoka, Y., Webster, R.G. and Paulson, J.C. (1994) Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology*, **205**, 17-23.
34. Cros, J.F., Garcia-Sastre, A. and Palese, P. (2005) An unconventional NLS is critical for the nuclear import of the influenza A virus nucleoprotein and ribonucleoprotein. *Traffic*, **6**, 205-213.
35. Desselberger, U., Racaniello, V.R., Zazra, J.J. and Palese, P. (1980) The 3' and 5'-terminal sequences of influenza A, B and C virus RNA segments are highly conserved and show partial inverted complementarity. *Gene*, **8**, 315-328.
36. Dias, A., Bouvier, D., Crepin, T., McCarthy, A.A., Hart, D.J., Baudin, F., Cusack, S.

and Ruigrok, R.W. (2009) The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature*, **458**, 914-918.

37. Digard, P., Elton, D., Bishop, K., Medcalf, E., Weeds, A. and Pope, B. (1999) Modulation of nuclear localization of the influenza virus nucleoprotein through interaction with actin filaments. *J Virol*, **73**, 2222-2231.
38. Doms, R.W., Lamb, R.A., Rose, J.K. and Helenius, A. (1993) Folding and assembly of viral membrane proteins. *Virology*, **193**, 545-562.
39. Duhaut, S.D. and McCauley, J.W. (1996) Defective RNAs inhibit the assembly of influenza virus genome segments in a segment-specific manner. *Virology*, **216**, 326-337.
40. Elton, D., Medcalf, L., Bishop, K., Harrison, D. and Digard, P. (1999) Identification of amino acid residues of influenza virus nucleoprotein essential for RNA binding. *J Virol*, **73**, 7357-7367.
41. Elton, D., Medcalf, E., Bishop, K. and Digard, P. (1999) Oligomerization of the influenza virus nucleoprotein: identification of positive and negative sequence elements. *Virology*, **260**, 190-200.
42. Elton, D., Simpson-Holley, M., Archer, K., Medcalf, L., Hallam, R., McCauley, J. and Digard, P. (2001) Interaction of the influenza virus nucleoprotein with the cellular CRM1-mediated nuclear export pathway. *J Virol*, **75**, 408-419.
43. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, **60**, 2126-2132.
44. Enami, M., Sharma, G., Benham, C. and Palese, P. (1991) An influenza virus containing nine different RNA segments. *Virology*, **185**, 291-298.
45. Epstein, S.L., Kong, W.P., Misplon, J.A., Lo, C.Y., Tumpey, T.M., Xu, L. and Nabel, G.J. (2005) Protection against multiple influenza A subtypes by vaccination with highly conserved nucleoprotein. *Vaccine*, **23**, 5404-5410.
46. Faure, G., Bornot, A. and de Brevern, A.G. (2008) Protein contacts, inter-residue interactions and side-chain modelling. *Biochimie*, **90**, 626-639.
47. Fernandez-Fuentes, N., Zhai, J. and Fiser, A. (2006) ArchPRED: a template based loop structure prediction server. *Nucleic Acids Res*, **34**, W173-176.
48. Fislova, T., Thomas, B., Graef, K.M. and Fodor, E. (2010) Association of the influenza virus RNA polymerase subunit PB2 with the host chaperonin CCT. *J Virol*, **84**, 8691-8699.
49. Flick, R., Neumann, G., Hoffmann, E., Neumeier, E. and Hobom, G. (1996) Promoter elements in the influenza vRNA terminal structure. *Rna*, **2**, 1046-1057.
50. Fodor, E., Pritlove, D.C. and Brownlee, G.G. (1994) The influenza virus panhandle is involved in the initiation of transcription. *J Virol*, **68**, 4092-4096.

51. Fodor, E., Crow, M., Mingay, L.J., Deng, T., Sharps, J., Fechter, P. and Brownlee, G.G. (2002) A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. *J Virol*, **76**, 8989-9001.
52. Fouchier, R.A., Munster, V., Wallensten, A., Bestebroer, T.M., Herfst, S., Smith, D., Rimmelzwaan, G.F., Olsen, B. and Osterhaus, A.D. (2005) Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. *J Virol*, **79**, 2814-2822.
53. Galegov, G.A., Andronova, V.L. and Nebol'sin, V.E. (2009) [Antiviral effect of Ingavirin against seasonal influenza virus A/H1N1 in MDCK cell culture]. *Antibiot Khimioter*, **54**, 19-22.
54. Gerhard, W., Mozdzanowska, K. and Zharikova, D. (2006) Prospects for universal influenza virus vaccine. *Emerg Infect Dis*, **12**, 569-574.
55. Gomez-Puertas, P., Albo, C., Perez-Pastrana, E., Vivo, A. and Portela, A. (2000) Influenza virus matrix protein is the major driving force in virus budding. *J Virol*, **74**, 11538-11547.
56. Gong, J., Fang, H., Li, M., Liu, Y., Yang, K. and Xu, W. (2009) Potential targets and their relevant inhibitors in anti-influenza fields. *Curr Med Chem*, **16**, 3716-3739.
57. Gonzalez, S., Zurcher, T. and Ortin, J. (1996) Identification of two separate domains in the influenza virus PB1 protein involved in the interaction with the PB2 and PA subunits: a model for the viral RNA polymerase structure. *Nucleic Acids Res*, **24**, 4456-4463.
58. Gonzalez, S. and Ortin, J. (1999) Characterization of influenza virus PB1 protein binding to viral RNA: two separate regions of the protein contribute to the interaction domain. *J Virol*, **73**, 631-637.
59. Gonzalez, S. and Ortin, J. (1999) Distinct regions of influenza virus PB1 polymerase subunit recognize vRNA and cRNA templates. *Embo J*, **18**, 3767-3775.
60. Green, T.J., Zhang, X., Wertz, G.W. and Luo, M. (2006) Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science*, **313**, 357-360.
61. Gschoesser, C., Almanzar, G., Hainz, U., Ortin, J., Schonitzer, D., Schild, H., Saurwein-Teissl, M. and Grubeck-Loebenstein, B. (2002) CD4+ and CD8+ mediated cellular immune response to recombinant influenza nucleoprotein. *Vaccine*, **20**, 3731-3738.
62. Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R.W., Ortin, J., Hart, D.J. and Cusack, S. (2008) The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat Struct*

Mol Biol, **15**, 500-506.

63. Hagiwara, K., Kondoh, Y., Ueda, A., Yamada, K., Goto, H., Watanabe, T., Nakata, T., Osada, H. and Aida, Y. (2010) Discovery of novel antiviral agents directed against the influenza A virus nucleoprotein using photo-cross-linked chemical arrays. *Biochem Biophys Res Commun*, **394**, 721-727.
64. Han, X., Bushweller, J.H., Cafiso, D.S. and Tamm, L.K. (2001) Membrane structure and fusion-triggering conformational change of the fusion domain from influenza hemagglutinin. *Nat Struct Biol*, **8**, 715-720.
65. Hatta, M., Gao, P., Halfmann, P. and Kawaoka, Y. (2001) Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. *Science*, **293**, 1840-1842.
66. Hay, A.J., Lomniczi, B., Bellamy, A.R. and Skehel, J.J. (1977) Transcription of the influenza virus genome. *Virology*, **83**, 337-355.
67. Hay, A.J., Skehel, J.J. and McCauley, J. (1982) Characterization of influenza virus RNA complete transcripts. *Virology*, **116**, 517-522.
68. He, X., Zhou, J., Bartlam, M., Zhang, R., Ma, J., Lou, Z., Li, X., Li, J., Joachimiak, A., Zeng, Z., Ge, R., Rao, Z. and Liu, Y. (2008) Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. *Nature*, **454**, 1123-1126.
69. Hirayama, E., Atagi, H., Hiraki, A. and Kim, J. (2004) Heat shock protein 70 is related to thermal inhibition of nuclear export of the influenza virus ribonucleoprotein complex. *J Virol*, **78**, 1263-1270.
70. Hsu, M.T., Parvin, J.D., Gupta, S., Krystal, M. and Palese, P. (1987) Genomic RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle. *Proc Natl Acad Sci U S A*, **84**, 8140-8144.
71. Huang, T.S., Palese, P. and Krystal, M. (1990) Determination of influenza virus proteins required for genome replication. *J Virol*, **64**, 5669-5673.
72. Janin, J., Bahadur, R.P. and Chakrabarti, P. (2008) Protein-protein interaction and quaternary structure. *Q Rev Biophys*, **41**, 133-180.
73. Johnson, N.P. and Mueller, J. (2002) Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic. *Bull Hist Med*, **76**, 105-115.
74. Jorba, N., Coloma, R. and Ortin, J. (2009) Genetic trans-complementation establishes a new model for influenza virus RNA transcription and replication. *PLoS Pathog*, **5**, e1000462.
75. Kao, R.Y., Yang, D., Lau, L.S., Tsui, W.H., Hu, L., Dai, J., Chan, M.P., Chan, C.M., Wang, P., Zheng, B.J., Sun, J., Huang, J.D., Madar, J., Chen, G., Chen, H., Guan, Y. and Yuen, K.Y. (2010) Identification of influenza A nucleoprotein as an antiviral

target. *Nat Biotechnol*, **28**, 600-605.

76. Kawaguchi, A. and Nagata, K. (2007) De novo replication of the influenza virus RNA genome is regulated by DNA replicative helicase, MCM. *Embo J*, **26**, 4566-4575.
77. Ketha, K.M. and Atreya, C.D. (2008) Application of bioinformatics-coupled experimental analysis reveals a new transport-competent nuclear localization signal in the nucleoprotein of influenza A virus strain. *BMC Cell Biol*, **9**, 22.
78. Kilbourne, E.D. (2006) Influenza pandemics of the 20th century. *Emerg Infect Dis*, **12**, 9-14.
79. Kim, H.J., Fodor, E., Brownlee, G.G. and Seong, B.L. (1997) Mutational analysis of the RNA-fork model of the influenza A virus vRNA promoter in vivo. *J Gen Virol*, **78 (Pt 2)**, 353-357.
80. Klenk, H.D. and Garten, W. (1994) Host cell proteases controlling virus pathogenicity. *Trends Microbiol*, **2**, 39-43.
81. Klumpp, K., Ruigrok, R.W. and Baudin, F. (1997) Roles of the influenza virus polymerase and nucleoprotein in forming a functional RNP structure. *Embo J*, **16**, 1248-1257.
82. Kobayashi, M., Toyoda, T., Adyshev, D.M., Azuma, Y. and Ishihama, A. (1994) Molecular dissection of influenza virus nucleoprotein: deletion mapping of the RNA binding domain. *J Virol*, **68**, 8433-8436.
83. Kreijtz, J.H., de Mutsert, G., van Baalen, C.A., Fouchier, R.A., Osterhaus, A.D. and Rimmelzwaan, G.F. (2008) Cross-recognition of avian H5N1 influenza virus by human cytotoxic T-lymphocyte populations directed to human influenza A virus. *J Virol*, **82**, 5161-5166.
84. Krug, R.M., Broni, B.A. and Bouloy, M. (1979) Are the 5' ends of influenza viral mRNAs synthesized in vivo donated by host mRNAs? *Cell*, **18**, 329-334.
85. Kundu, A., Avalos, R.T., Sanderson, C.M. and Nayak, D.P. (1996) Transmembrane domain of influenza virus neuraminidase, a type II protein, possesses an apical sorting signal in polarized MDCK cells. *J Virol*, **70**, 6508-6515.
86. Kuzuhara, T., Kise, D., Yoshida, H., Horita, T., Murazaki, Y., Nishimura, A., Echigo, N., Utsunomiya, H. and Tsuge, H. (2009) Structural basis of the influenza A virus RNA polymerase PB2 RNA-binding domain containing the pathogenicity-determinant lysine 627 residue. *J Biol Chem*, **284**, 6855-6860.
87. Labadie, K., Dos Santos Afonso, E., Rameix-Welti, M.A., van der Werf, S. and Naffakh, N. (2007) Host-range determinants on the PB2 protein of influenza A viruses control the interaction between the viral polymerase and nucleoprotein in human cells. *Virology*, **362**, 271-282.

88. Lamb, R.A. and R.M. Krug. (2001) Orhomyxoviridae: the viruses and their replication. In Fields Virology. Vol. 1. Fourth edition D.M. Knipe and P.M. Howley, Eds.: 1487-1531. Lippincott Williams and Wilkins. Philadelphia.

89. Laver, W.G., Colman, P.M., Webster, R.G., Hinshaw, V.S. and Air, G.M. (1984) Influenza virus neuraminidase with hemagglutinin activity. *Virology*, **137**, 314-323.

90. Leahy, M.B., Dobbyn, H.C. and Brownlee, G.G. (2001) Hairpin loop structure in the 3' arm of the influenza A virus virion RNA promoter is required for endonuclease activity. *J Virol*, **75**, 7042-7049.

91. Li, M.L., Ramirez, B.C. and Krug, R.M. (1998) RNA-dependent activation of primer RNA production by influenza virus polymerase: different regions of the same protein subunit constitute the two required RNA-binding sites. *Embo J*, **17**, 5844-5852.

92. Li, M.L., Rao, P. and Krug, R.M. (2001) The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits. *Embo J*, **20**, 2078-2086.

93. Li, Z., Watanabe, T., Hatta, M., Watanabe, S., Nanbo, A., Ozawa, M., Kakugawa, S., Shimojima, M., Yamada, S., Neumann, G. and Kawaoka, Y. (2009) Mutational analysis of conserved amino acids in the influenza A virus nucleoprotein. *J Virol*, **83**, 4153-4162.

94. Li, O.T., Chan, M.C., Leung, C.S., Chan, R.W., Guan, Y., Nicholls, J.M. and Poon, L.L. (2009) Full factorial analysis of mammalian and avian influenza polymerase subunits suggests a role of an efficient polymerase for virus adaptation. *PLoS One*, **4**, e5658.

95. Lin, S., Naim, H.Y., Rodriguez, A.C. and Roth, M.G. (1998) Mutations in the middle of the transmembrane domain reverse the polarity of transport of the influenza virus hemagglutinin in MDCK epithelial cells. *J Cell Biol*, **142**, 51-57.

96. Liu, T., Muller, J. and Ye, Z. (2002) Association of influenza virus matrix protein with ribonucleoproteins may control viral growth and morphology. *Virology*, **304**, 89-96.

97. Livingston, B.D., Higgins, D. and Van Nest, G. (2006) Evolving strategies for the prevention of influenza infection: potential for multistrain targeting. *BioDrugs*, **20**, 335-340.

98. Luo, M., Green, T.J., Zhang, X., Tsao, J. and Qiu, S. (2007) Structural comparisons of the nucleoprotein from three negative strand RNA virus families. *Virol J*, **4**, 72.

99. Marcotrigiano, J., Gingras, A.C., Sonenberg, N. and Burley, S.K. (1997) Cocrystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to

7-methyl-GDP. *Cell*, **89**, 951-961.

100. Martin, K. and Helenius, A. (1991) Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (M1) promotes export and inhibits import. *Cell*, **67**, 117-130.
101. Martin-Benito, J., Area, E., Ortega, J., Llorca, O., Valpuesta, J.M., Carrascosa, J.L. and Ortin, J. (2001) Three-dimensional reconstruction of a recombinant influenza virus ribonucleoprotein particle. *EMBO Rep*, **2**, 313-317.
102. Mase, M., Tanimura, N., Imada, T., Okamatsu, M., Tsukamoto, K. and Yamaguchi, S. (2006) Recent H5N1 avian influenza A virus increases rapidly in virulence to mice after a single passage in mice. *J Gen Virol*, **87**, 3655-3659.
103. Massin, P., van der Werf, S. and Naffakh, N. (2001) Residue 627 of PB2 is a determinant of cold sensitivity in RNA replication of avian influenza viruses. *J Virol*, **75**, 5398-5404.
104. McCullers, J.A., Saito, T. and Iverson, A.R. (2004) Multiple genotypes of influenza B virus circulated between 1979 and 2003. *J Virol*, **78**, 12817-12828.
105. Medcalf, L., Poole, E., Elton, D. and Digard, P. (1999) Temperature-sensitive lesions in two influenza A viruses defective for replicative transcription disrupt RNA binding by the nucleoprotein. *J Virol*, **73**, 7349-7356.
106. Mehle, A. and Doudna, J.A. (2008) An inhibitory activity in human cells restricts the function of an avian-like influenza virus polymerase. *Cell Host Microbe*, **4**, 111-122.
107. Mehle, A. and Doudna, J.A. (2009) Adaptive strategies of the influenza virus polymerase for replication in humans. *Proc Natl Acad Sci U S A*, **106**, 21312-21316.
108. Mena, I., Jambrina, E., Albo, C., Perales, B., Ortin, J., Arrese, M., Vallejo, D. and Portela, A. (1999) Mutational analysis of influenza A virus nucleoprotein: identification of mutations that affect RNA replication. *J Virol*, **73**, 1186-1194.
109. Minor, W., Cymborowski, M., Otwinowski, Z. and Chruszcz, M. (2006) HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr*, **62**, 859-866.
110. Miotto, O., Heiny, A., Tan, T.W., August, J.T. and Brusic, V. (2008) Identification of human-to-human transmissibility factors in PB2 proteins of influenza A by large-scale mutual information analysis. *BMC Bioinformatics*, **9 Suppl 1**, S18.
111. Momose, F., Basler, C.F., O'Neill, R.E., Iwamatsu, A., Palese, P. and Nagata, K. (2001) Cellular splicing factor RAF-2p48/NPI-5/BAT1/UAP56 interacts with the influenza virus nucleoprotein and enhances viral RNA synthesis. *J Virol*, **75**, 1899-1908.

112. Momose, F., Naito, T., Yano, K., Sugimoto, S., Morikawa, Y. and Nagata, K. (2002) Identification of Hsp90 as a stimulatory host factor involved in influenza virus RNA synthesis. *J Biol Chem*, **277**, 45306-45314.
113. Moncorge, O., Mura, M. and Barclay, W.S. (2010) Evidence for avian and human host cell factors that affect the activity of influenza virus polymerase. *J Virol*, **84**, 9978-9986.
114. Mukaigawa, J. and Nayak, D.P. (1991) Two signals mediate nuclear localization of influenza virus (A/WSN/33) polymerase basic protein 2. *J Virol*, **65**, 245-253.
115. Muller, R., Poch, O., Delarue, M., Bishop, D.H. and Bouloy, M. (1994) Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. *J Gen Virol*, **75 (Pt 6)**, 1345-1352.
116. Murphy, B.R., Hinshaw, V.S., Sly, D.L., London, W.T., Hosier, N.T., Wood, F.T., Webster, R.G. and Chanock, R.M. (1982) Virulence of avian influenza A viruses for squirrel monkeys. *Infect Immun*, **37**, 1119-1126.
117. Naffakh, N., Massin, P., Escriou, N., Crescenzo-Chaigne, B. and van der Werf, S. (2000) Genetic analysis of the compatibility between polymerase proteins from human and avian strains of influenza A viruses. *J Gen Virol*, **81**, 1283-1291.
118. Naffakh, N., Massin, P. and van der Werf, S. (2001) The transcription/replication activity of the polymerase of influenza A viruses is not correlated with the level of proteolysis induced by the PA subunit. *Virology*, **285**, 244-252.
119. Naito, T., Momose, F., Kawaguchi, A. and Nagata, K. (2007) Involvement of Hsp90 in assembly and nuclear import of influenza virus RNA polymerase subunits. *J Virol*, **81**, 1339-1349.
120. Nath, S.T. and Nayak, D.P. (1990) Function of two discrete regions is required for nuclear localization of polymerase basic protein 1 of A/WSN/33 influenza virus (H1 N1). *Mol Cell Biol*, **10**, 4139-4145.
121. Neumann, G., Castrucci, M.R. and Kawaoka, Y. (1997) Nuclear import and export of influenza virus nucleoprotein. *J Virol*, **71**, 9690-9700.
122. Neumann, G., Hughes, M.T. and Kawaoka, Y. (2000) Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. *Embo J*, **19**, 6751-6758.
123. Ng, A.K., Zhang, H., Tan, K., Li, Z., Liu, J.H., Chan, P.K., Li, S.M., Chan, W.Y., Au, S.W., Joachimiak, A., Walz, T., Wang, J.H. and Shaw, P.C. (2008) Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *Faseb J*, **22**, 3638-3647.

124. Ng, A.K., Wang, J.H. and Shaw, P.C. (2009) Structure and sequence analysis of influenza A virus nucleoprotein. *Sci China C Life Sci*, **52**, 439-449.
125. Nieto, A., de la Luna, S., Barcena, J., Portela, A. and Ortin, J. (1994) Complex structure of the nuclear translocation signal of influenza virus polymerase PA subunit. *J Gen Virol*, **75 (Pt 1)**, 29-36.
126. Noda, T. and Kawaoka, Y. Structure of influenza virus ribonucleoprotein complexes and their packaging into virions. *Rev Med Virol*, **20**, 380-391.
127. Noda, T., Sagara, H., Yen, A., Takada, A., Kida, H., Cheng, R.H. and Kawaoka, Y. (2006) Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature*, **439**, 490-492.
128. Normile, D. (2006) Avian influenza. WHO proposes plan to stop pandemic in its tracks. *Science*, **311**, 315-316.
129. O'Neill, R.E., Jaskunas, R., Blobel, G., Palese, P. and Moroianu, J. (1995) Nuclear import of influenza virus RNA can be mediated by viral nucleoprotein and transport factors required for protein import. *J Biol Chem*, **270**, 22701-22704.
130. Obayashi, E., Yoshida, H., Kawai, F., Shibayama, N., Kawaguchi, A., Nagata, K., Tame, J.R. and Park, S.Y. (2008) The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature*, **454**, 1127-1131.
131. Odagiri, T. and Tashiro, M. (1997) Segment-specific noncoding sequences of the influenza virus genome RNA are involved in the specific competition between defective interfering RNA and its progenitor RNA segment at the virion assembly step. *J Virol*, **71**, 2138-2145.
132. Ohi, M., Li, Y., Cheng, Y. and Walz, T. (2004) Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biol Proced Online*, **6**, 23-34.
133. Ohtsu, Y., Honda, Y., Sakata, Y., Kato, H. and Toyoda, T. (2002) Fine mapping of the subunit binding sites of influenza virus RNA polymerase. *Microbiol Immunol*, **46**, 167-175.
134. Ortega, J., Martin-Benito, J., Zurcher, T., Valpuesta, J.M., Carrascosa, J.L. and Ortin, J. (2000) Ultrastructural and functional analyses of recombinant influenza virus ribonucleoproteins suggest dimerization of nucleoprotein during virus amplification. *J Virol*, **74**, 156-163.
135. Ozawa, M., Fujii, K., Muramoto, Y., Yamada, S., Yamayoshi, S., Takada, A., Goto, H., Horimoto, T. and Kawaoka, Y. (2007) Contributions of two nuclear localization signals of influenza A virus nucleoprotein to viral replication. *J Virol*, **81**, 30-41.
136. Palese, P., Tobita, K., Ueda, M. and Compans, R.W. (1974) Characterization of temperature sensitive influenza virus mutants defective in neuraminidase.

Virology, **61**, 397-410.

137. Palese, P. and Compans, R.W. (1976) Inhibition of influenza virus replication in tissue culture by 2-deoxy-2,3-dehydro-N-trifluoroacetylneuraminic acid (FANA): mechanism of action. *J Gen Virol*, **33**, 159-163.
138. Palese, P. (2004) Influenza: old and new threats. *Nat Med*, **10**, S82-87.
139. Perez, J.T., Varble, A., Sachidanandam, R., Zlatev, I., Manoharan, M., Garcia-Sastre, A. and tenOever, B.R. (2010) Influenza A virus-generated small RNAs regulate the switch from transcription to replication. *Proc Natl Acad Sci U S A*, **107**, 11525-11530.
140. Plotch, S.J., Bouloy, M., Ulmanen, I. and Krug, R.M. (1981) A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell*, **23**, 847-858.
141. Poch, O., Sauvaget, I., Delarue, M. and Tordo, N. (1989) Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *Embo J*, **8**, 3867-3874.
142. Pons, M.W., Schulze, I.T., Hirst, G.K. and Hauser, R. (1969) Isolation and characterization of the ribonucleoprotein of influenza virus. *Virology*, **39**, 250-259.
143. Poole, E., Elton, D., Medcalf, L. and Digard, P. (2004) Functional domains of the influenza A virus PB2 protein: identification of NP- and PB1-binding sites. *Virology*, **321**, 120-133.
144. Poon, L.L., Pritlove, D.C., Fodor, E. and Brownlee, G.G. (1999) Direct evidence that the poly(A) tail of influenza A virus mRNA is synthesized by reiterative copying of a U track in the virion RNA template. *J Virol*, **73**, 3473-3476.
145. Poon, L.L., Fodor, E. and Brownlee, G.G. (2000) Polyuridylated mRNA synthesized by a recombinant influenza virus is defective in nuclear export. *J Virol*, **74**, 418-427.
146. Portela, A. and Digard, P. (2002) The influenza virus nucleoprotein: a multifunctional RNA-binding protein pivotal to virus replication. *J Gen Virol*, **83**, 723-734.
147. Pritlove, D.C., Poon, L.L., Devenish, L.J., Leahy, M.B. and Brownlee, G.G. (1999) A hairpin loop at the 5' end of influenza A virus virion RNA is required for synthesis of poly(A)+ mRNA in vitro. *J Virol*, **73**, 2109-2114.
148. Prokudina, E.N., Semenova, N.P., Chumakov, V.M. and Rudneva, I.A. (2004) Transient disulfide bonds formation in conformational maturation of influenza virus nucleocapsid protein (NP). *Virus Res*, **99**, 169-175.
149. Prokudina-Kantorovich, E.N. and Semenova, N.P. (1996) Intracellular

oligomerization of influenza virus nucleoprotein. *Virology*, **223**, 51-56.

150. Rameix-Welti, M.A., Tomoiu, A., Dos Santos Afonso, E., van der Werf, S. and Naffakh, N. (2009) Avian Influenza A virus polymerase association with nucleoprotein, but not polymerase assembly, is impaired in human cells during the course of infection. *J Virol*, **83**, 1320-1331.
151. Richardson, J.C. and Akkina, R.K. (1991) NS2 protein of influenza virus is found in purified virus and phosphorylated in infected cells. *Arch Virol*, **116**, 69-80.
152. Rimmelzwaan, G.F., Boon, A.C., Voeten, J.T., Berkhoff, E.G., Fouchier, R.A. and Osterhaus, A.D. (2004) Sequence variation in the influenza A virus nucleoprotein associated with escape from cytotoxic T lymphocytes. *Virus Res*, **103**, 97-100.
153. Robertson, J.S. (1979) 5' and 3' terminal nucleotide sequences of the RNA genome segments of influenza virus. *Nucleic Acids Res*, **6**, 3745-3757.
154. Robertson, J.S., Schubert, M. and Lazzarini, R.A. (1981) Polyadenylation sites for influenza virus mRNA. *J Virol*, **38**, 157-163.
155. Roose, K., Fiers, W. and Saelens, X. (2009) Pandemic preparedness: toward a universal influenza vaccine. *Drug News Perspect*, **22**, 80-92.
156. Rota, P.A., Hemphill, M.L., Whistler, T., Regnery, H.L. and Kendal, A.P. (1992) Antigenic and genetic characterization of the haemagglutinins of recent cocirculating strains of influenza B virus. *J Gen Virol*, **73 (Pt 10)**, 2737-2742.
157. Roti, M., Yang, J., Berger, D., Huston, L., James, E.A. and Kwok, W.W. (2008) Healthy human subjects have CD4+ T cells directed against H5N1 influenza virus. *J Immunol*, **180**, 1758-1768.
158. Ruigrok, R.W. and Baudin, F. (1995) Structure of influenza virus ribonucleoprotein particles. II. Purified RNA-free influenza virus ribonucleoprotein forms structures that are indistinguishable from the intact influenza virus ribonucleoprotein particles. *J Gen Virol*, **76 (Pt 4)**, 1009-1014.
159. Sanz-Ezquerro, J.J., de la Luna, S., Ortin, J. and Nieto, A. (1995) Individual expression of influenza virus PA protein induces degradation of coexpressed proteins. *J Virol*, **69**, 2420-2426.
160. Shapiro, G.I. and Krug, R.M. (1988) Influenza virus RNA replication in vitro: synthesis of viral template RNAs and virion RNAs in the absence of an added primer. *J Virol*, **62**, 2285-2290.
161. Shu, L.L., Bean, W.J. and Webster, R.G. (1993) Analysis of the evolution and variation of the human influenza A virus nucleoprotein gene from 1933 to 1990. *J Virol*, **67**, 2723-2729.
162. Skehel, J.J. and Hay, A.J. (1978) Nucleotide sequences at the 5' termini of influenza virus RNAs and their transcripts. *Nucleic Acids Res*, **5**, 1207-1219.

163. Skehel, J.J., Bayley, P.M., Brown, E.B., Martin, S.R., Waterfield, M.D., White, J.M., Wilson, I.A. and Wiley, D.C. (1982) Changes in the conformation of influenza virus hemagglutinin at the pH optimum of virus-mediated membrane fusion. *Proc Natl Acad Sci U S A*, **79**, 968-972.
164. Skorko, R., Summers, D.F. and Galarza, J.M. (1991) Influenza A virus in vitro transcription: roles of NS1 and NP proteins in regulating RNA synthesis. *Virology*, **180**, 668-677.
165. Steinhauer, D.A. (1999) Role of hemagglutinin cleavage for the pathogenicity of influenza virus. *Virology*, **258**, 1-20.
166. Su, C.Y., Cheng, T.J., Lin, M.I., Wang, S.Y., Huang, W.I., Lin-Chu, S.Y., Chen, Y.H., Wu, C.Y., Lai, M.M., Cheng, W.C., Wu, Y.T., Tsai, M.D., Cheng, Y.S. and Wong, C.H. (2010) High-throughput identification of compounds targeting influenza RNA-dependent RNA polymerase activity. *Proc Natl Acad Sci U S A*, **107**, 19151-19156.
167. Sugiyama, K., Obayashi, E., Kawaguchi, A., Suzuki, Y., Tame, J.R., Nagata, K. and Park, S.Y. (2009) Structural insight into the essential PB1-PB2 subunit contact of the influenza virus RNA polymerase. *Embo J*, **28**, 1803-1811.
168. Tarendeau, F., Boudet, J., Guilligay, D., Mas, P.J., Bougault, C.M., Boulo, S., Baudin, F., Ruigrok, R.W., Daigle, N., Ellenberg, J., Cusack, S., Simorre, J.P. and Hart, D.J. (2007) Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nat Struct Mol Biol*, **14**, 229-233.
169. Tarendeau, F., Crepin, T., Guilligay, D., Ruigrok, R.W., Cusack, S. and Hart, D.J. (2008) Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog*, **4**, e1000136.
170. Tiley, L.S., Hagen, M., Matthews, J.T. and Krystal, M. (1994) Sequence-specific binding of the influenza virus RNA polymerase to sequences located at the 5' ends of the viral RNAs. *J Virol*, **68**, 5108-5116.
171. Tovchigrechko, A. and Vakser, I.A. (2006) GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res*, **34**, W310-314.
172. Ulmanen, I., Broni, B.A. and Krug, R.M. (1981) Role of two of the influenza virus core P proteins in recognizing cap 1 structures (m7GpppNm) on RNAs and in initiating viral RNA transcription. *Proc Natl Acad Sci U S A*, **78**, 7355-7359.
173. Vreede, F.T., Jung, T.E. and Brownlee, G.G. (2004) Model suggesting that replication of influenza virus is regulated by stabilization of replicative intermediates. *J Virol*, **78**, 9568-9572.

174. Wang, P., Palese, P. and O'Neill, R.E. (1997) The NPI-1/NPI-3 (karyopherin alpha) binding site on the influenza a virus nucleoprotein NP is a nonconventional nuclear localization signal. *J Virol*, **71**, 1850-1856.
175. Wang, P., Song, W., Mok, B.W., Zhao, P., Qin, K., Lai, A., Smith, G.J., Zhang, J., Lin, T., Guan, Y. and Chen, H. (2009) Nuclear factor 90 negatively regulates influenza virus replication by interacting with viral nucleoprotein. *J Virol*, **83**, 7850-7861.
176. Weber, F., Kochs, G., Gruber, S. and Haller, O. (1998) A classical bipartite nuclear localization signal on Thogoto and influenza A virus nucleoproteins. *Virology*, **250**, 9-18.
177. Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R.F., Sykes, B.D. and Wishart, D.S. (2003) VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res*, **31**, 3316-3319.
178. World Health Organization (2010) Pandemic (H1N1) 2009 – update 112. Available at http://www.who.int/csr/don/2010_08_06/en/index.html
179. World Health Organization (2011) Cumulative Number of Confirmed Human Cases of Avian Influenza A/(H5N1) Reported to WHO. Available at http://www.who.int/csr/disease/avian_influenza/country/cases_table_2011_04_11/en/index.html
180. Yager, E.J., Ahmed, M., Lanzer, K., Randall, T.D., Woodland, D.L. and Blackman, M.A. (2008) Age-associated decline in T cell repertoire diversity leads to holes in the repertoire and impaired immunity to influenza virus. *J Exp Med*, **205**, 711-723.
181. Yamanaka, K., Ishihama, A. and Nagata, K. (1990) Reconstitution of influenza virus RNA-nucleoprotein complexes structurally resembling native viral ribonucleoprotein cores. *J Biol Chem*, **265**, 11151-11155.
182. Yasuda, J., Nakada, S., Kato, A., Toyoda, T. and Ishihama, A. (1993) Molecular assembly of influenza virus: association of the NS2 protein with virion matrix. *Virology*, **196**, 249-255.
183. Ye, Z., Liu, T., Offringa, D.P., McInnis, J. and Levandowski, R.A. (1999) Association of influenza virus matrix protein with ribonucleoproteins. *J Virol*, **73**, 7467-7473.
184. Ye, Q., Krug, R.M. and Tao, Y.J. (2006) The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature*, **444**, 1078-1082.
185. Yuan, P., Bartlam, M., Lou, Z., Chen, S., Zhou, J., He, X., Lv, Z., Ge, R., Li, X., Deng, T., Fodor, E., Rao, Z. and Liu, Y. (2009) Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature*, **458**, 909-913.

186. Yuen, K.Y., Chan, P.K., Peiris, M., Tsang, D.N., Que, T.L., Shortridge, K.F., Cheung, P.T., To, W.K., Ho, E.T., Sung, R. and Cheng, A.F. (1998) Clinical features and rapid viral diagnosis of human disease associated with avian influenza A H5N1 virus. *Lancet*, **351**, 467-471.
187. Zhang, K. and Sun, F. (2005) Assessing the power of tag SNPs in the mapping of quantitative trait loci (QTL) with extremal and random samples. *BMC Genet*, **6**, 51.
188. Zheng, H., Lee, H.A., Palese, P. and Garcia-Sastre, A. (1999) Influenza A virus RNA polymerase has the ability to stutter at the polyadenylation site of a viral RNA template during RNA replication. *J Virol*, **73**, 5240-5243.