Title

Structural basis for RNA-binding and homo-oligomer formation by influenza B virus nucleoprotein

Running Title

Influenza B NP Structure

Authors and Affiliations

Andy Ka-Leung Ng1,*, Mandy Ka-Han Lam1,*, Hongmin Zhang2,^, Jinhuan Liu2, Shannon Wing-Ngor Au1, Paul Kay-Sheung Chan3, Jiahuai Wang2,# and Pang-Chui Shaw1,#

1 Centre for Protein Science and Crystallography, School of Life Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong, China
2 Department of Medical Oncology, Dana-Farber Cancer Institute; Department of Pediatrics, and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA
3 Department of Microbiology, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, Hong Kong, China

Copyright © 2012, American Society for Microbiology. All Rights Reserved.
* These authors contribute equally in this work.

^ Present address: Department of Physiology, The University of Hong Kong, Pokfulam, Hong Kong, China

# Corresponding authors

Pang-Chui Shaw: pcshaw@cuhk.edu.hk

Jiahuai Wang: jwang@red.dfc.harvard.edu
Abstract

Influenza virus nucleoprotein (NP) is the major component of the viral ribonucleoprotein complex, which is crucial for the transcription and replication of the viral genome. We have determined the crystal structure of influenza B virus NP to a resolution of 3.2 Å. Influenza B NP contains a head, a body domain and a tail loop. The electropositive groove between the head and body domains of influenza B NP is crucial for RNA binding. This groove also contains an extended flexible charged loop (aa. 125-149) and two lysine clusters at the first half of this loop were shown to be crucial for binding RNA. Influenza B NP forms a crystallographic homo-tetramer by inserting the tail loop into the body domain of the neighboring NP molecule. A deeply buried salt bridge R472-E395 and a hydrophobic cluster at F468 are the major driving forces for the insertion. The analysis of the influenza B virus NP structure and function and comparisons with influenza A virus NP provide insights into the mechanisms of action and underpin efforts to design inhibitors for this class of proteins.
Introduction

Influenza viruses are RNA viruses and classified into three types: A, B and C. While much attention has been paid to influenza A virus, the severity of influenza B virus cannot be underestimated. In the past 70 years, there have been 16 epidemics, resulting in excess morbidity and mortality, at least partially caused by influenza B virus (34). From July 2010 to June 2011, 25.0% of influenza positive specimens found globally were of influenza B virus (33). Influenza B virus also causes substantial mortality among pediatric patients. Among the 116 deaths associated with influenza infections occurring in the 2010-2011 flu season in the United States, 45 were due to influenza B virus (6). Therefore, it is important to find out how influenza B virus functions and compare it with influenza A virus, so that more effective means could be developed to combat the highly infectious influenza viruses in general.

The genome of influenza B virus comprises eight negative-sense RNA segments encoding eleven polypeptides (14). Among these proteins, nucleoprotein (NP) is the major component of the ribonucleoprotein complex (RNP), which consists of RNA, NP and RNA polymerase and plays a vital role in transcription and replication of
the viral genome (reviewed in ref. 27).

Influenza B NP (BNP) is a basic protein (pI>9) with 560 amino acids and molecular weight of 62 kDa. The primary sequence of BNP has several distinctive features compared to influenza A NP (ANP): (1) BNP contains a significantly extended N-terminal region (aa. 1-70); (2) both nuclear localization signal (NLS)-1 and NLS-2 in ANP (31, 32) are absent in BNP; (3) some known RNA-binding regions in ANP, especially the flexible basic loop of aa 74-88 (21), are not conserved in BNP; (4) the linkers and tail loops for NP homo-oligomerization are not conserved; (5) gaps are found when aligning the C-terminus of ANP and BNP. Recently, the crystal structures of ANP from H5N1 and H1N1 viruses have been determined by us and others (12, 21, 35). Here we report the crystal structure of BNP (PDB code: 3TJ0) at 3.2Å and compare its structure and function with those of ANP for RNA binding and for forming oligomers.
Materials and Methods

Biological materials

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). Anti-BNP antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA), anti-myc antibody (Cell Signaling Technology), anti-flag antibody (Sigma-Aldrich, Louis, MO, USA) and anti-beta-actin antibody (GenScript, Piscataway, NJ, USA) were purchased commercially. Plasmids pCIPA, pCIPB1 and pCIPB2 expressing RNA polymerase subunits of influenza B/Panama/45/90 virus were described previously (15). Plasmid pPol-Luc-RT for the generation of pPol-Luci-BNA-RT and pEGFP were kindly provided by L.L.M. Poon of the University of Hong Kong (18). The pPol-Luci-BNA-RT transcribes a vRNA-like RNA, in which the non-coding sequences of influenza B NA segment flank the coding region of firefly luciferase. Plasmid pcDNA-BNP was generated by inserting the wild-type and mutant NP gene of B/HongKong/CUHK-24964/2004 into mammalian expression vector pcDNA3 (Invitrogen) for BNP expression in 293T cells. The genes of wild-type and mutant BNP were also cloned into pcDNA™3.1/myc-His (Invitrogen) for myc-tagged BNP expression in mammalian cells. pCMV-Tag2B obtained from Agilent Technologies, Inc.,
Santa Clara, CA, USA, was for flag-tagged BNP expression in mammalian cells, and pRHisMBP obtained from K.B. Wong, the Chinese University of Hong Kong, was for the expression of maltose binding protein (MBP)-tagged BNP variants in Escherichia coli.

Expression and Purification of influenza B NP

MBP-tagged BNP variants were expressed in Escherichia coli C41 (DE3). The cells were lysed in 20 mM sodium phosphate, 150 mM NaCl, pH 6.5. The lysate was passed through an amylose column (New England Biolabs, Ipswich, MA, USA). The bound protein was eluted in buffer containing 20 mM maltose. The eluate was incubated with thrombin (100 U) (Sigma-Aldrich) and RNase A (300 U) (Sigma-Aldrich) at 4°C overnight to remove MBP tag and RNA from BNP. It was then passed through a heparin high-performance (HP) column (GE Healthcare, Waukesha, WI, USA). NP was eluted with a 0 to 1.5 M NaCl gradient in the same buffer. For crystallization purpose, the BNP protein was concentrated and buffer exchanged into 20 mM MOPS, 100 mM NaCl, pH 7.0.

Crystallization and Structure Determination

RNase-treated NP in a concentration of 15 mg/ml, was mixed in a 1:1 ratio with 1.4
8 M phosphate, pH 7.0. Crystals were grown in hanging drops at 16 °C. Crystals were frozen in crystallization buffer with 20 % glycerol, and brought to Shanghai Synchrotron Radiation Facility for data collection. The data were collected at 100 K using beamline BL17U (with X-rays at a wavelength of 0.97941 Å) and were processed and scaled with the HKL2000 suite (22). The structure was determined by molecular replacement using the tail loop-deleted influenza A H5N1 NP monomer (21) as the search model. Calculations and model building were carried out using Phenix (1) and Coot (10) respectively. The final model was analyzed by the program PROCHECK (16). Figures of protein structures were prepared with the program PyMOL (9).

Polymerase activity assay

Plasmids pCIPA, pCIPB1, pCIPB2, pcDNA-BNP (wild-type or NP mutants) and pPol-Luci-BNA-RT (0.125 μg each, except the amount of pcDNA-NP was individually adjusted according to their expression levels) were co-transfected to 1.125x10^5 293T cells for RNP complex reconstitution. Plasmid for EGFP expression (0.06 μg) was also co-transfected for normalization (18). For the negative control, empty pcDNA plasmid instead of pcDNA-BNP was transfected to cells. At 48 h post-transfection, 293T cells were lysed by Steady-Glo assay reagent (Promega) for five minutes...
followed by luminescence measurement on a luminometer (Victor2 1420 Multilabel Counter, Wallac) according to manufacturer’s instructions.

**Static Light Scattering**

Wild-type or variant NP proteins were subject to static light scattering using a miniDAWN triangle (45°, 90°, and 135°) light scattering detector (Wyatt Technology, Santa Barbara, CA) connected to an Optilab DSP interferometric refractometer (Wyatt Technology Corporation, Santa Barbara, CA, USA). This system was connected to a Superdex 200 column (GE Healthcare) controlled by an AKTAexplorer chromatography system (GE Healthcare). Before sample injection, the miniDAWN detector system was equilibrated with 100 mM sodium phosphate (pH 6.0) and 100 mM NaCl for at least 2 h to ensure a stable baseline signal. The flow rate was set to 0.5 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. Wyatt ASTRA software was used to evaluate all data obtained.

**Co-immunoprecipitation**

Two μg of flag-tagged and myc-tagged NP plasmids were transfected into human kidney 293T cells in suspension. Since the presence of RNA may cause non-specific
interaction of BNP molecules without the tail loop insertion event, the G1 cluster in some of these myc- and flag- tagged BNP constructs were mutated to alanine, for preventing the binding of RNA from interfering with the results interpretation. Co-immunoprecipitation was performed at 48 h posttransfection. Cells were lysed in 50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, and 1% Triton X-100, pH 7.6 (co-IP buffer). The lysate was centrifuged at 16,000 X g for 10 min at 4°C. The supernatant was treated with 150 U RNaseA (Sigma-Aldrich), and incubated at 4°C overnight with or without anti-Myc antibody (Cell Signaling Technology, Danvers, MA, USA). The mixture was then incubated with protein A beads for 1.5 h at 4°C with shaking. The beads were centrifuged and washed with co-IP buffer three times before being boiled in SDS loading dye and analyzed by Western blotting.

Surface Plasmon Resonance

A biotinylated 2′-O-methylated RNA oligonucleotide with the sequence 5′-UUU GUU ACA CAC ACA CAC GCU GUG-3′ was prepared in running buffer (10 mM Tris-HCl, 150 mM NaCl, 0.005 % surfactant P20, 1 mM DTT, 5 % glycerol, pH 8) and immobilized on an SA sensor chip (GE Healthcare) until the surface density reached 30 to 35 response units (RU), according to manufacturer’s instructions (GE Healthcare). Kinetic measurements were carried out with a BIAcore 3000 system at 25°C.
Concentration series of NP variants were injected onto the SA chip at 30 μl/min in running buffer. 2M NaCl was used to regenerate the chip surface. Data were analyzed with BIAevaluation v. 4.1 software, using the ‘1:1 Langmuir’ model.
Results

Crystal structure of influenza B Nucleoprotein

BNP was over-expressed in Escherichia coli and purified to more than 95% homogeneity. BNP formed oval-shaped crystals in the orthorhombic space group I222. A diffraction data set of 3.2 Å resolution was collected and the structure of BNP was determined by molecular replacement, using the influenza A H5N1 NP monomer (PDB: 2Q06, chain A) as the search model, with the linkers and tail loop region (aa. 395-437) excluded. The structure was refined to $R/ R_{\text{free}} = 0.245/0.291$ (Table 1). One asymmetric unit contains two BNP molecules, named chains A and B (Figure 1A). In the structure, 458 residues (aa. 73-126, 147-550) were modeled for chain A while 457 residues (aa. 72-124, 148-551) for chain B. BNP was crystallized in tetrameric form in this study (Figure 1B), in contrast to the trimeric forms of ANP (12, 21, 35). Their differences and the biological implications of these will be further discussed below.

BNP folds into a head and a body domain both of which are largely helical. The polypeptide chain goes back and forth several times between the two domains, which could impose a constraint on the relative movement of the two domains.
The electron densities for the first 71 residues of BNP were not visible, indicating that this N-terminal region of the protein is highly flexible. The head domain consists of aa. 212-323 and 496-509, while the body domain is formed by aa. 72-208, 336-404 and 521-551. Like what is seen in ANP, a special feature of BNP is a long tail loop (aa. 459-486) which inserts into a neighboring molecule, facilitating the formation of BNP homo-oligomers (Figure 1C). These structural features are linked together by several loop regions (aa. 209-211, 324-335, 405-458, 487-495 and 510-520) (Figure 1D).

The overall folds of influenza A and B NP are similar, yet a few regions have high root-mean-square deviation (RMSD) of Cα and some regions could not be aligned together. For example, the tail loop is oriented differently in the two structures. The least-aligned regions are mostly flexible surface loops. One of these loops (aa. 263-271) in BNP was found to be a cytotoxic T lymphocytes epitope (28).

The role of the basic groove and the flexible charged loop in RNA binding

Purified influenza virus NP could bind single-stranded RNA, but with no or little sequence specificity (4). Purified monomeric ANP mutants were found to have significantly reduced RNA-binding affinity. On the other hand, purified trimeric and
tetrameric ANP, which better represent the oligomeric NP in physiological states,
bear high affinity towards RNA (5, 7).

In the BNP structure, an electropositive groove is found between the head and body
domains of BNP which is likely to interact with RNA (Figure 2A). Two clusters of
positively charged residues are identified in this groove, including clusters G1 (aa.
K125, K126, R235, R236) and G2 (aa. R211, K213, R217). Charge-to-alanine
mutations on these two clusters reduced the polymerase activity to less than 10% of
the wild-type (Table 2). Surface plasmon resonance (SPR) also showed that the
protein variants were unable to interact with RNA (Table 2). It is noted that the
homo-oligomerization of the variants was not affected by the mutations (Figure 2B).
The findings are consistent with the phenotypes of the corresponding ANP mutants
(21, data not shown).

Sequence alignment between ANP and BNP however identifies a much extended
flexible charged loop in BNP between aa. 125 and aa. 149. It contains nine basic
and six acidic residues (Figure 2C). Only aa. 125-126 and 147-149 were visible in
the density map, showing that this loop is highly flexible. Deletion of this loop in
BNP resulted in a 14-fold decrease in the RNA-binding affinity, which is significantly
more dramatic than the 6.4-fold decrease in the flexible loop-deleted ANP equivalent
(21), implying that this loop in BNP contributes more to RNA-binding. While the
loop-deleted BNP mutant rendered the polymerase inactive, mutation of the
negative charges in the loop (D/E-to-A) did not cause significant reduction in
polymerase activity (Table 2). However, when the positively-charged residues in the
loop were mutated to alanine, the K/R-to-A mutant lost most of the polymerase
activity (Table 2). SPR data of the K/R-to-A mutant also showed no RNA binding,
with affinity lower than that of the loop-deleted mutant (Table 2). This is possibly
due to the charge-charge repulsion of the remaining acidic residues in the loop
against the negatively charged RNA immobilized on the chip. Nevertheless, the
positive charges in the loop are essential for RNA-binding affinity, as well as
polymerase activity in the context of viral RNP. To further map the crucial basic
residues on the flexible charged loop, we constructed another four multiple-point
charged-to-alanine mutants. Residues R136, K139, K142 and K149 at the second
half of the loop were found to be non-essential, as mutations of these did not reduce
polymerase activity (Table 2). In contrast, five-point mutations on the first half of
the loop (K125A, K126A, K131A, K132A and K133A forming two clusters) reduced
polymerase activity to 5.85% of the wild-type; and the protein mutant was unable to
bind RNA in the SPR measurement (Table 2). However, mutating individual clusters
([K125A, K126A] or [K131A, K132A, K133A]) did not cause reduction in polymerase activities (Table 2). These data indicate that the two lysine clusters are important to the viral RNP activity and the either lysine cluster is sufficient for the flexible basic loop to bind RNA.

BNP homo-oligomerizes by tail loop insertion

In the tetrameric crystal structure of BNP, each tetramer is composed of two A chains and two B chains (Figure 1B). This is largely facilitated by the tail loop insertion of one molecule into the next, following the order: from chain A to chain B', to chain A'', to chain B''', back to chain A. The use of such an extended loop near the C-terminus for homo-oligomerization is a common strategy adopted by NP molecules of different viruses, including influenza A virus (12, 21, 35), rabies virus (2) and vesicular stomatitis virus (13). The BNP tail loop (aa. 458-484) (Figure 3A) inserts deeply into the body groove of a neighboring NP molecule. The insertion is crucial for the function of the RNP, as deletion of the tail loop resulted in a dramatic reduction (3.52% of the wild-type) in polymerase activity (Table 3). Static light scattering showed that the tail loop-deleted BNP exists mainly in monomers, in contrast to the tetramer population of wild-type BNP in solution (Figure 4A). The small proportion of dimer observed is probably due to the relatively weak interaction between the body
domains of two BNP molecules, as seen in the crystal asymmetric unit (Figure 1A).

As co-immunoprecipitation showed that myc- and flag- tagged tail loop deleted BNP

do not interact with each other (Figure 4C), it is the tail loop insertion that provides
the major driving force for oligomerization.

The crystal structure of BNP reveals an extensive network of intra-chain (clusters T1
to T3) and inter-chain (clusters I1 to I6) interactions at the tail loop region, which we
classified into different clusters according to their proximity and nature of
interactions (Figures 3A and 3B). To understand the role of these clusters towards
BNP homo-oligomerization, the involved hydrophilic and hydrophobic residues were
mutated to alanine and serine, respectively, for the elimination of the interactions
concerned. The expression levels of these mutants were normalized (data not
shown) before investigating their effect on polymerase activity and oligomeric states.

Although most of the tail loop interactions are hydrophobic in nature, a salt bridge
between E395 and R472 (cluster I2) is found deeply buried within (Figure 3C).

Mutating either residue to alanine led to dramatic decrease in polymerase activities
(4.1 % and 7.6 % of the wild-type, respectively) (Table 3). Both the E395A and
R472A protein mutants resulted in the loss of homo-oligomerization ability in vitro
The main chains of F468 and V470 also form hydrogen bonds with the side chain of R472 (cluster T1), which possibly helps to orient R472 for the formation of inter-chain salt bridge with E395.

The tip region of the tail loop not only comprises a salt bridge, but also contains an extensive network of hydrophobic contacts (cluster I1). The tail loop residue F468 interacts with M360, F390, V399, L400 and L403 of the neighboring BNP (Figure 3D). Mutation of F468S led to a dramatic decrease in polymerase activity (3.1 % of the wild-type). Static light scattering and co-immunoprecipitation also demonstrated that the F468S mutant could not form homo-oligomers (Figures 4B and 4C). Mutating F412 (the BNP F468 equivalent) in ANP also had similar effects (8, 17). Although this residue is conserved in ANP and BNP, there are significant differences in its interactions (Figure 3D). Only one out of the five residues that interact with it is homologous (V399 in BNP and I/L/V343 in influenza A NP). Even though the interactions are conserved, F390 and L403 in BNP become H/N334 and I/V347 in ANP, respectively. The ANP equivalents of BNP, M360 and L400, do not play any role in the interaction.

The hydrophobic residues in the stem of the BNP tail loop also interact with the
neighboring BNP molecule (Figure 3E), including residues in cluster I3 (P473, I474 and L476) and cluster I4 (C464, P466 and V467). Single- and double-point mutations in these clusters did not cause a major effect on polymerase activities (Table 3). However, when all three residues were mutated to serine (Myc-tagged [P473S, I474S, L476S] and [C464S, P466S, V467S]), polymerase activities dropped significantly (2.8 % and 8.9 % of the wild-type) (Table 3). These two mutants could not form BNP oligomers either (Figure 4B and 4C). No major structural changes were detected for these two triple-point mutants using circular dichroism spectroscopy (data not shown). Many of the contacts in clusters I3 and I4 are non-conserved between ANP and BNP (Figure 3E). Structural alignment between the tail loops (not the full length protein) of ANP and BNP (Figure 3F) shows that the rmsd of residues P473 and I474 is higher than that of the rest of the loop (Figure 3G). Residue P473 in BNP (N417 in ANP) restricts the phi-psi dihedral angles. The conserved I474 (L418 in ANP) faces the neighboring BNP molecule instead of having van der Waal’s contacts with the conserved P466 (P410 in ANP) (Figure 3F). The crucial zig-zag pattern of hydrophobic interactions at the stem for maintaining the tail loop conformation in ANP (7) is thus not observed in BNP.

The base of the tail loop features clusters T3, I5 and I6. While cluster T3 is not
essential, a hydrogen bond between R325 and C464 (cluster I5) plays a certain role in tail loop insertion, as the polymerase activity of the R325A mutant dropped to 41% of the wild-type (Table 3). Regarding cluster I6, a similar interaction pair is found in ANP (I406-Y487) and mutation of I406S renders the polymerase partially active (7). A BNP I462S mutation did not cause major reduction in polymerase activity (Table 3), which showed that cluster I6 is likely to have limited functional importance.
Influenza B NP is indispensable for the transcription and replication of the influenza B virus RNA genome. This report reveals the structure of influenza B virus NP and analyses the functional importance of amino acid residues involved in RNA binding and homo-oligomerization.

So far, all NP crystal structures from single-stranded negative sense RNA viruses have a two-domain fold (2, 11, 13, 23, 24, 27, 30, 35). For influenza B NP, the proposed RNA-binding groove is found at the periphery of the tetrameric BNP (Figure 5A), which is similar to the docked 9-mer NPA model after EM 3D reconstruction (8). On the contrary, the RNA of rabies and vesicular stomatitis viruses is completely sequestered in the NP polymer ring and the opening to access the RNA faces the center of the ring, as revealed in their crystal structures (2, 13). Besides, one NP molecule of rabies and vesicular stomatitis viruses bind 9 nucleotides of RNA (2, 13) while one NP molecule of influenza A virus binds 24-27 RNA nucleotides in vivo (3, 19).

Not only NP of single-stranded negative sense RNA viruses encapsidates RNA in
different manners, they also adopt various strategies to form homo-oligomers. Lassa virus NP uses its N-terminal domain to interact with the C-terminal domain of the neighboring molecule (23). The extreme N-terminus of rift-valley fever virus NP forms an alpha helical arm for inserting into the neighboring molecule (11). Human respiratory syncytial virus and borna disease virus NP use their extreme N- and C-termini to interact with the neighboring NP, but without tight contacts (27, 30). Rabies and vesicular stomatitis viruses NP use both an extended loop and the extreme N-terminus for the interactions (2, 13). While influenza NPs also utilize the extended tail loop for homo-oligomer formation, whether the N-terminal region takes part in the process like the other viral NPs is yet to be determined.

Nine ANP monomers are organized in a ring-like structure in a reconstituted mini-RNP (19). Based on the flexibility between the tail loop and the rest of ANP structure observed, we have attempted to construct a 9-mer model (21). The newly determined BNP structure has provided some support to this model. Despite the fact that different interactions at the central helices (α11, α16 and α17) are found between BNP tetramer and ANP trimer, several hydrophobic interactions are likely to be maintained. Residue L485 in BNP helix α16 interacts with the neighboring molecule through residue L316 in helix α11 and residues L501 and M504.
of in helix α17 (Figure 5B, ii, iii). Accordingly, the equivalent ANP residue F429 interacts with F258, I445 and M448 (Figure 5B, i, iv). These conserved interactions in ANP trimer and BNP tetramer are likely to be kept in NP higher oligomers. The loop region in helix-loop-helix motif (aa. 458-509), for connecting the tail loop to the main body of NP, may vary in structural organization upon the formation of different oligomers, as we have observed that A494 and D495 of BNP are part of the loop while their corresponding residues in ANP S438 and D439 are part of the helix (Figure 5C). Therefore, the structural features observed for the formation of BNP tetramer, together with the trimeric ANP, provides insights into how NP forms higher oligomers.

The N-terminal regions of both ANP and BNP are not visible in the crystal structure. BNP, however, has a much extended N-terminal region than ANP, as revealed by their sequence comparisons. This N-terminus is non-conserved among the >300 BNP sequences in the NCBI database. 23 out of the 70 residues (32.9 %) are polymorphic (defined as bearing more than one amino acid variation in that particular residue), and the percentage of polymorphism is much higher than of the rest of the BNP (8.8 %). While the N-terminus of ANP has been well-characterized for carrying an unconventional nuclear localization signal (31) and a binding site for
the cellular splicing factor BAT1/UAP56 (20), no particular functions were found to associate with this region of BNP (29).

Structural comparisons of ANP and BNP tail loop interactions reveal a strikingly conserved R472-E395 salt bridge (R416-E339 in ANP). Primary sequence analysis also showed that these two residues are conserved among different strains of influenza A virus and different lineages of influenza B virus, which indicate the essentiality of the interaction. The charged side-chains of the two residues neutralize one another and the salt bridge is deeply buried in the hydrophobic insertion groove. The identification of ligands which prevent the formation of the salt bridge will be of particular interest to combat both types of influenza viruses. Intriguingly, an inhibitor which targets the salt bridge of ANP has been recently identified (28). It is interesting to find if this inhibitor also targets the BNP salt bridge.

Through the study of the structure and functional analysis of BNP, we have identified crucial residues for RNA-binding and tail loop insertion. The only conserved inter-chain interaction of R472-E395 in influenza A and B NP strengthens the idea that ligands targeting this salt bridge would be a universal drug for combating both A
Comparison of the influenza A NP trimer and BNP tetramer structures suggest that the structural reorganization of the helix-loop-helix motif and the maintenance of some conserved hydrophobic contacts in higher-order NP oligomer formation.
Acknowledgments

The authors thank Professor Wendy Barclay at Imperial College, University of London for providing pCIPA, pCIPB1 and pCIPB2 plasmids and Professor Ervin Fodor at University of Oxford for critical reading of the manuscript. This research was supported by a General Research Fund grant (CUHK473810) from the Research Grants Council of Hong Kong (PCS) and the Claudia Adams Barr Program in Cancer Research (JW). Equipment used was obtained through a Special Equipment Grant (SEG CUHK08) from the University Grants Council of Hong Kong. The diffraction data were collected at the Shanghai Synchrotron Radiation Facility (Shanghai, China).


Figure 1. Crystal structure of BNP. (A) Two molecules in the asymmetric unit of the BNP crystal structure. (B) BNP forms a crystallographic tetramer, with the tail loop inserting in the order of A-B'-A''-B'''. (C) BNP (surface representation in blue) inserts its tail loop into the neighboring molecule (pink). (D) Structural organization of BNP (chain A). \(\alpha\)-helices and \(\beta\)-strands are numbered. Anti-parallel \(\beta\)-strands are denoted with ‘–’.

Figure 2. RNA-binding of BNP. (A) Electrostatic potential representation of BNP. The groove between the head and body domains is highly electropositive. Residues in the G1 and G2 clusters are shown as stick models. The estimated position of the flexible charged loop is indicated in circle with broken lines. (B) Co-immunoprecipitation of BNP RNA-binding mutants. The three protein variants which showed no binding towards RNA in the SPR experiment could homo-oligomerize. The DNA sequences of the mutants were confirmed by sequencing. (C) Sequence alignment of influenza A and B NP shows that BNP has a much lengthened charged loop (underlined) while G1 (■) and G2 (▲) clusters are mostly conserved.
Figure 3. Tail loop interaction of BNP. (A) Intra-chain interactions of the tail loop. Electrostatic and hydrophobic interactions are in red and black dotted lines respectively. (B) Inter-chain interactions between the tail loop and the neighboring BNP. Tail loop residues are in pale green boxes while residues in neighboring BNP are in purple boxes. (C) The E395-R472 salt bridge between the tail loop (green) and the neighboring NP (purple). The position of the R472 side chain is oriented by the hydrogen bond interaction with the main chain O of F468 and V470. (D) BNP F468 and ANP F412 make distinct interactions with the neighboring molecule. Residues in sticks interact with F468/F412 while residues in lines do not. M360, L400 and M445 of BNP interact with F468, but the equivalent F304, S344 and R38 of ANP do not. (E) Residues C464, P466, V467, P473, I474 and L476 in the stem region of the BNP tail loop (green) make novel contacts with the neighboring molecule (pink). These interactions are not observed in ANP. The stem region of the ANP tail loop forms zig-zag pattern of intra-chain hydrophobic interaction instead. (F) Structural alignment of the tail loops in ANP (grey) and BNP (green). I474 of BNP faces towards the neighboring molecule but not towards P466, which is different from L418 in ANP. (G) Graph of rmsd of individual residues when the stem and tip regions of the tail loops in ANP and BNP are aligned. Residues P473 and I474 are
Figure 4. Static light scattering and co-immunoprecipitation of the homo-oligomerization mutants. (A) Static light scattering of wild-type and tail loop-deleted BNP. Curves are refractive interference signal. Values are native molecular weights of BNP in the population of the peaks. (B) Static light scattering of the polymerase activity defective BNP mutants, compared to the wild-type. (C) Co-immunoprecipitation of BNP oligomerization mutants. NP without RNA-binding activity (WT/G1) could self-interact and served as the positive control. All BNP oligomerization mutants with defective polymerase activity could not homo-oligomerize.

Figure 5. Tetramer formation of BNP. (A) Electrostatic potential representation of BNP tetramer. RNA-binding groove is located at the periphery of the tetramer, at the highly positively charged region. An obvious space is observed in the middle of the tetramer. (B) Interactions of the central helices of ANP and BNP. (i) Hydrophobic interactions between α11 and α18 in ANP. (ii) In BNP, M484 of the tail loop forms electrostatic interactions with R319 of the neighboring BNP molecule. V481 of the tail loop also forms hydrophobic interactions with V322 and V323 of the
neighboring BNP. M484 and L485 form another van der Waal’s contacts with R319.

(iii) K478 and N505 do not form salt bridges in BNP. Instead, M508, which is not found in influenza A NP, interacts with K478, V481 and R482. (iv) The R422-E449 salt bridge is prominent in ANP. (C) Sequence alignment of influenza A and B shows that the regions for homo-oligomerization are less conserved (29 % identical).

Shaded regions are the helices α16 and α17 of the helix-loop-helix motif. α17 is especially mis-aligned between the two NPs and suggests that the motif is flexible in nature.
<table>
<thead>
<tr>
<th></th>
<th>BNP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>I 2 2 2</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>(a, b, c) (Å)</td>
<td>106.95, 123.34, 198.08</td>
</tr>
<tr>
<td>(\alpha, \beta, \gamma) (°)</td>
<td>90.00, 90.00, 90.00</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>104.73-3.23 (3.35-3.23) *</td>
</tr>
<tr>
<td>(R_{sym}) or (R_{merge})</td>
<td>0.113 (0.387)</td>
</tr>
<tr>
<td>(I / \sigma I)</td>
<td>13.8 (2.9)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>93.2 (95.1)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>7.2 (7.3)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>47.055-3.233</td>
</tr>
<tr>
<td>No. reflections</td>
<td>19862</td>
</tr>
<tr>
<td>(R_{work}) / (R_{free})</td>
<td>0.245 / 0.291</td>
</tr>
<tr>
<td>No. atoms</td>
<td>7022</td>
</tr>
<tr>
<td>(B)-factors</td>
<td>77.10</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.002</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>0.729</td>
</tr>
<tr>
<td><strong>Ramachandran plot</strong></td>
<td></td>
</tr>
<tr>
<td>Most favoured</td>
<td>88.5 %</td>
</tr>
<tr>
<td>Additional allowed</td>
<td>9.0 %</td>
</tr>
<tr>
<td>Generously allowed</td>
<td>2.5 %</td>
</tr>
<tr>
<td>Disallowed</td>
<td>0.0 %</td>
</tr>
</tbody>
</table>

*Data from a single crystal.

*Values in parentheses are for highest-resolution shell.
<table>
<thead>
<tr>
<th>Target region</th>
<th>Influenza B NP</th>
<th>Relative luciferase activity at 48 h post-transfection (%)</th>
<th>P</th>
<th>Mean K&lt;sub&gt;D&lt;/sub&gt; (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/</td>
<td>Wild type</td>
<td>99.00 ± 3.42</td>
<td>1.28 ± 0.3</td>
<td></td>
</tr>
<tr>
<td>/</td>
<td>Negative control&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.56 ± 0.04</td>
<td>**&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>K125A, K126A, R235A, R236A</td>
<td>2.06 ± 0.58</td>
<td>**&lt;sup&gt;e&lt;/sup&gt;</td>
<td>ND</td>
</tr>
<tr>
<td>G2</td>
<td>R211A, K213A, R217A</td>
<td>2.58 ± 0.44</td>
<td>**&lt;sup&gt;e&lt;/sup&gt;</td>
<td>ND</td>
</tr>
<tr>
<td>125&lt;sup&gt;KKTEFQKKNA&lt;/sup&gt;NDVKEGKEEIDHNK&lt;sup&gt;(49)&lt;/sup&gt;</td>
<td>ΔΔ25-146</td>
<td>2.14 ± 0.60</td>
<td>**&lt;sup&gt;e&lt;/sup&gt;</td>
<td>195 ± 9.5</td>
</tr>
<tr>
<td>125&lt;sup&gt;KKTA&lt;/sup&gt;FDQKVKNKARVKA&lt;sup&gt;AT&lt;/sup&gt;A&lt;sup&gt;AH&lt;/sup&gt;NNK&lt;sup&gt;(49)&lt;/sup&gt;</td>
<td>D/E-to-A</td>
<td>47.77 ± 3.23</td>
<td>**&lt;sup&gt;e&lt;/sup&gt;</td>
<td>--</td>
</tr>
<tr>
<td>125&lt;sup&gt;AA&lt;/sup&gt;TEFQAAA&lt;sup&gt;AD&lt;/sup&gt;VAE&lt;sup&gt;GD&lt;/sup&gt;EEIDHNK&lt;sup&gt;(49)&lt;/sup&gt;</td>
<td>K/R-to-A</td>
<td>2.92 ± 1.39</td>
<td>**&lt;sup&gt;e&lt;/sup&gt;</td>
<td>ND</td>
</tr>
<tr>
<td>125&lt;sup&gt;KKTEFQK&lt;sup&gt;KN&lt;/sup&gt;ADVKE&lt;sup&gt;KA&lt;/sup&gt;GKEEIDHNK&lt;sup&gt;(49)&lt;/sup&gt;</td>
<td>R136A, K139A, K142A, K149A</td>
<td>108.21 ± 1.58</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>125&lt;sup&gt;AA&lt;/sup&gt;TEFQAAA&lt;sup&gt;AD&lt;/sup&gt;VAE&lt;sup&gt;GD&lt;/sup&gt;EEIDHNK&lt;sup&gt;(49)&lt;/sup&gt;</td>
<td>K125A, K126A, K131A, K132A, K133A</td>
<td>5.85 ± 0.84</td>
<td>**&lt;sup&gt;e&lt;/sup&gt;</td>
<td>--</td>
</tr>
<tr>
<td>125&lt;sup&gt;KKTEFQK&lt;sup&gt;KN&lt;/sup&gt;ADVKE&lt;sup&gt;KA&lt;/sup&gt;GKEEIDHNK&lt;sup&gt;(49)&lt;/sup&gt;</td>
<td>K125A, K126A</td>
<td>101.17 ± 3.46</td>
<td>*</td>
<td>--</td>
</tr>
</tbody>
</table>

<sup>a</sup> The residues with charge-to-alanine mutation are underlined. The relative luciferase activity reflects the ability of NP mutants to support polymerase activity. Mutants which exhibited <10 % activity of the wild-type are defined as having very low or no polymerase activity. Mutants with 11-50 % activity are defined as partially active. Mutants whose activities are 51-100 % are active. (*, P < 0.01; **, P < 0.001)

<sup>b</sup> The dissociation constant K<sub>D</sub> was measured by SPR. N-terminal (aa. 1-66)-deleted BNP variants were used to prevent unwanted interference of the chip surface. It is noted that the N-terminus did not interact with RNA (data not shown) and it has been previously shown that the N-terminus is not required for the expression and replication of a model RNA [25]. ND, the RNA-binding affinity is too weak to determined; --, not attempted.

<sup>c</sup> Transfection was carried out without the BNP plasmid.
### Table 3  Polymerase activity of BNP homo-oligomerization mutants

<table>
<thead>
<tr>
<th>Cluster</th>
<th>NP mutant disrupting the interactions in the cluster</th>
<th>Relative luciferase activity at 48 h post-transfection (%)</th>
<th>P&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>99.00 ± 3.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative control</td>
<td>0.56 ± 0.04</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>Δtail loop (aa. 458-484)</td>
<td>3.52 ± 0.91</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>Myc-tagged wild type</td>
<td>100.00 ± 12.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myc-tagged Δtail loop</td>
<td>4.44 ± 0.86</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>R472A&lt;sup&gt;b&lt;/sup&gt;</td>
<td>7.66 ± 1.34</td>
<td>**</td>
</tr>
<tr>
<td>T2</td>
<td>C464S&lt;sup&gt;b&lt;/sup&gt;</td>
<td>120.11 ± 7.90</td>
<td>*</td>
</tr>
<tr>
<td>A475S</td>
<td>124.98 ± 18.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>I462S&lt;sup&gt;b&lt;/sup&gt;</td>
<td>71.73 ± 1.79</td>
<td>*</td>
</tr>
<tr>
<td>L476S&lt;sup&gt;b&lt;/sup&gt;</td>
<td>77.77 ± 9.22</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>A480S</td>
<td>108.35 ± 5.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I1</td>
<td>F468S</td>
<td>3.07 ± 1.49</td>
<td>**</td>
</tr>
<tr>
<td>I2</td>
<td>E395A</td>
<td>4.14 ± 1.05</td>
<td>**</td>
</tr>
<tr>
<td>R472A&lt;sup&gt;b&lt;/sup&gt;</td>
<td>7.66 ± 1.34</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>I3</td>
<td>P473S</td>
<td>125.84 ± 23.03</td>
<td></td>
</tr>
<tr>
<td>I474S</td>
<td>106.02 ± 2.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L476S&lt;sup&gt;b&lt;/sup&gt;</td>
<td>77.77 ± 9.22</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>[P473S,I474S]</td>
<td>27.90 ± 4.91</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>[P473S,L476S]</td>
<td>49.11 ± 3.08</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>[I474S,L476S]</td>
<td>20.77 ± 2.41</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>Myc-tagged [P473S,I474S,L476S]&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2.78 ± 1.39</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>I4</td>
<td>C464S&lt;sup&gt;b&lt;/sup&gt;</td>
<td>120.11 ± 7.90</td>
<td>*</td>
</tr>
<tr>
<td>P466S</td>
<td>79.13 ± 6.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V467S</td>
<td>113.45 ± 8.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[464S,P466S]</td>
<td>53.71 ± 3.85</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>[C464S,V467S]</td>
<td>87.04 ± 2.71</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>[P466S,V467S]</td>
<td>63.79 ± 1.64</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>[C464S,P466S,V467S]</td>
<td>8.87± 1.24</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>I5</td>
<td>R325A</td>
<td>41.32 ± 0.79</td>
<td>**</td>
</tr>
<tr>
<td>I6</td>
<td>I462S&lt;sup&gt;b&lt;/sup&gt;</td>
<td>71.73 ± 1.79</td>
<td>*</td>
</tr>
</tbody>
</table>

<sup>a</sup> Mutants which could not be detected with the anti-BNP commercial antibodies were also cloned into myc-tagged vector, and their polymerase activities were compared to the myc-tagged wild-type NP.

<sup>b</sup> Mutants which are bolded appear in more than one cluster.

<sup>c</sup> *, P < 0.01; **, P < 0.001.