BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 7(2)

Antigenic variability in Neuraminidase protein of Influenza A/H3N2 vaccine strains (1968 – 2009)

Pratip Shil*, Sameer S Chavan, Sarah S Cherian

Bioinformatics and Data Management Division, National Institute of Virology, 20A Dr. Ambedkar Road, Pune- 411001, India; Pratip Shil – Email: shilpratip@gmail.com, pshil@icmr.org.in; Phone: +91 20 26006218; Fax: +91 20 26122669; *Corresponding author

Received July 23, 2011; Accepted July 25, 2011; Published September 06, 2011

Abstract:
Antigenic drift and shift involving the surface proteins of Influenza virus gave rise to new strains that caused epidemics affecting millions of people worldwide over the last hundred years. Variations in the membrane proteins like Hemagglutinin (HA) and Neuraminidase (NA) necessitates new vaccine strains to be updated frequently and poses challenge to effective vaccine design. Though the HA protein, the primary target of the human immune system, has been well studied, reports on the antigenic variability in the other membrane protein NA are sparse. In this paper we investigate the molecular basis of antigenic drift in the NA protein of the Influenza A/H3N2 vaccine strains between 1968 and 2009 and proceed to establish correlation between antigenic drift and antigen-antibody interactions. Sequence alignments and phylogenetic analyses were carried out and the antigenic variability was evaluated in terms of antigenic distance. To study the effects of antigenic drift on the protein structures, 3D structure of NA from various strains were predicted. Also, rigid body docking protocol has been used to study the interactions between these NA proteins and antibody Mem5, a 1998 antibody.

Keywords: Influenza virus, H3N2, neuraminidase, antibody, molecular docking, antigenic drift, host-pathogen interaction, Mem5.

Background:
Influenza is a major cause for concern worldwide due to the human suffering and economic burden afflicted by the periodic epidemics. The occurrence of repeated epidemics is attributed to effective evasion of herd immunity in host populations by the Influenza A viruses through antigenic variation in the surface proteins. Influenza A viruses have two surface glycoproteins– hemaglutinin (HA) and neuraminidase (NA). The antigenic variations in HA and NA proteins necessitate new vaccine strains to be updated frequently and hence constitutes a major challenge to effective vaccine design **[1]**. Though the antigenic variability of the HA protein has been studied **[2-5]**, reports related to antigenic variability of NA protein are rare. This may be due the fact that the NA protein, by the virtue of its positioning on the virus membrane, has somewhat limited interactions with the host immune system **[6]**. The Neuraminidase is a mushroom-shaped tetrameric protein, anchored to the viral membrane by a single sequence of 29 amino acid length at the N-terminus **[7]**. The four identical subunits are arranged with circular four fold symmetry and the enzyme active site is located centrally on each subunit. It is evident from reports that the anti-NA antibodies do provide protection and arguments have been made that NA should be included in the vaccines **[8, 9]**. However, direct demonstration of a significant contribution of NA to antigenic drift in human influenza viruses has not been made so far. There are instances of outbreaks due to failure of HA based vaccine. The 2003-2004 epidemic of H3N2 was caused by the proliferation of a new H3N2 subtype strain A/Fujian/411/2002 (an antigenic drift mutant of the A/Panama/2007/99 strain), with the existing vaccine failing to offer any protection **[1, 10]**. Hence, studies on antigenic variability in the NA protein are necessitated.

Though antigenic evolution of H3N2 has been studied in terms of HA protein, to the best of our knowledge reports involving NA protein of H3N2 viruses are sparse. In our earlier

publication **[11]**, we have reported about the antigenic variability in H3N2 virus strains (1968-2007) in the light of antigen-antibody interactions involving HA proteins. In the present paper, we studied the antigenic variability of NA proteins of influenza A/H3N2 (vaccine strains 1968-2009) using bioinformatics tools for sequence and structural analysis. Using modelling and docking protocols we also investigated the molecular basis of the reported (experimental evidence) failure of antibody Mem5 (raised against A/Memphis/31/98 NA protein and specific to epitope B) to bind to H3N2 strains evolving in 1999 and thereafter **[12]**.

Methodology:

The dataset:

The NA amino acid sequences for the following WHO recommended vaccine strains of influenza A/H3N2 virus were downloaded from the SwissProt database: A/Aichi/2/1968 (X31), A/Albany/1/1970 (ALB70), A/Port Chalmers/1/1973 (PC73), A/Memphis/31/98 (MP98), A/Moscow/10/1999 (MOS99), A/Wisconsin/67/2005 (WIS05), A/Brisbane/10/2007 (BR07) and A/Perth/16/2009 (PT09). The names in parenthesis will be used henceforth in the present paper. The structural information for Neuraminidase protein was retrieved from the Protein Databank (PDB ID: 2AEP). This known structure of MP98 NA protein complexed with antibody Mem5 has been considered as a reference for the study.

Phylogenetic analyses:

Sequence-based phylogenetic analyses were carried out using the Molecular Evolutionary Genetic Analyses (MEGA 4.0) package **[13]**. Multiple sequence alignment was carried out using the ClustalW implementation in MEGA 4.0 with default parameters. The phylogenetic tree was constructed using the neighbor-joining algorithm and bootstrap with 10,000 replications was used as test of phylogeny. To obtain the pairwise comparison of sequences (all possible combinations) within the dataset, the ALIGN algorithm as implemented in ISHAN package **[14]** has been used.

Antigenic variability:

In silico prediction of antigenic determinants was performed for each sequence of the dataset using the Kolaskar method **[15]** as implemented in the B-cell epitope prediction server at Immune Epitope Database (IEDB; www.immuneepitope.org/). The known antigenic regions **[15]** were also compared with the predictions. The antigenic distance between any two strains (newer vs older) can be measured in terms of the fraction of amino acids differing between the strains in the epitope regions. Such a measure is defined by $p_{epitope}$ **[1]**. It is calculated as follows: $p_{epitope}$ = Number of amino acid differences in the epitope/Total number of amino acids in the epitope

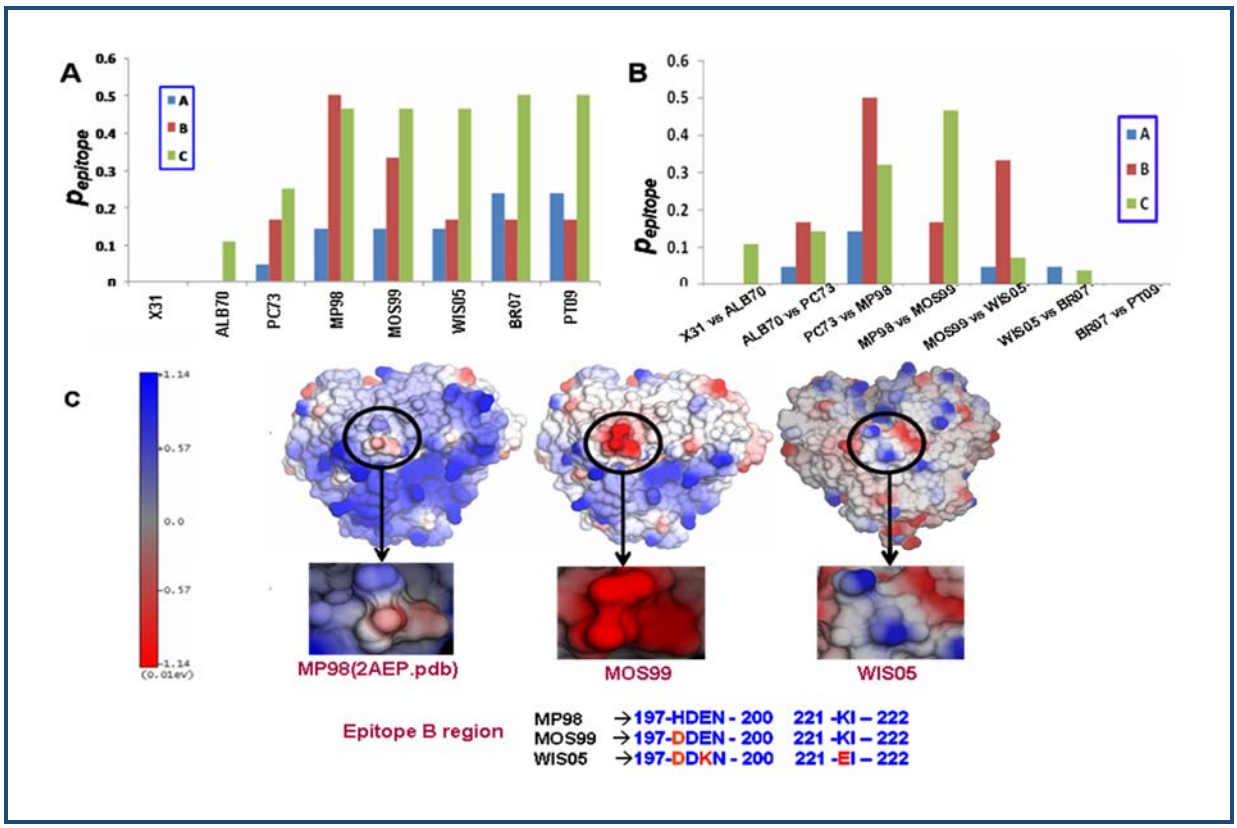

Figure 1: (A) Antigenic distances of all the epitopes (A, B and C) on NA proteins from selected vaccine strains of Influenza A/H3N2 computed with X31 as a standard. **(B)** Antigenic distances of all the epitopes (A, B and C) on NA proteins between successive strains of Influenza A/H3N2. **(C)** The comparison of surface electrostatics (outputs from NOC program) reveals dissimilarities between the MP98, MOS99 and WIS05 arising due to mutations in the epitope B. Numbering of amino acids are in accordance with sequence numbering of X31 (sequence ID: Q75VQ4).

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 7(2): 76-81 (2011)

3D structure prediction and analyses:
The 3D structures of the NA proteins from strains MOS99 and WIS05 have been predicted using the SwissModel online workstation. The template chosen (based on the automatic template selection mode) was the NA protein of MP98 (PDB ID: 2AEP.pdb). Energy minimization of the modeled structures and structural comparisons were performed using the GROMOS96 force field application in Swiss PDB-Viewer (SPDBV) **[16]**. Predicted 3D structures were evaluated using PROCHECK analyses **[17]**. Visualization of molecular structures and rendering of images was carried out in Discovery Studio v.2.0 (Accelyrs Inc., USA). The NOC software was used for studying the surface electrostatics of the proteins structures **[18]**.

Molecular docking:
Computational rigid-body docking protocol, a method which predicts the preferred stable orientation of one molecule when bound to a second one, was used to determine the binding site and the residues interacting between the antigens and antibody through hydrogen bonds (H-bonds) and salt bridges. *In silico* docking of antibody Mem5 against the NA proteins of MOS99 and WIS05 has been carried out using the ZDOCK server with default parameters **[19]**. No preconditions such as, specific amino acids desired to be at the interface, were specified at the time of running the docking protocol. From ten solutions returned by ZDOCK in each case, the best solution was selected based on the following conditions: i) the complimentarity determining regions (CDRs) of the antibody interacting with the antigen at the antigen-antibody interface and ii) the value of minimized energy of the complex being the least. The details of contacts between amino acids of antigen and antibody for each complex were obtained by Contacts of Structural Units (CSU) analyses **[20]**.

Discussion:
Neuraminidase amino acid sequences from various strains were subjected to pairwise comparison using the ISHAN package **[14]**. Identity (%) values calculated from these alignments were tabulated by the program **(Figure S1, see Supplementary material)**. Maximum identity (99.4%) in amino acid composition was observed between the strains BR07 and PT09. The minimum identity (85.5%) was observed between X31 and PT09. The net divergence within NA sequences considered in the study (time span: 1968-2009) was found to be 8.1%. Also, analyses based on earlier reports **[11]**, revealed that divergence between the HA sequences was 8.6% over the similar time span. Phylogenetic analyses further indicated clustering of the sequences and placed both BR07 and PT09 in one clade **(Figure S2, see Supplementary material)**. The multiple sequence alignment indicated the occurrence of mutations in the various strains with respect to X31. Compared to the X31 there were 16, 27, 53, 56, 63, 66 and 67 mutations in the sequences ALB70, PC73, MP98, MOS99, WIS05, BR07 and PT09 respectively **(Figure S3, see Supplementary material)**. The number of mutations between BR07 and PT09 were found to be only 3. These mutations were not in the epitopes or their vicinity. The amino acid stretches which recorded maximum variations in composition are: i) 140-155 and ii) epitope C covering 329-347 & 357-370 (as per numbering of complete amino acid sequences in **Figure S3, see Supplementary material**).

The antigenic evolution has been evaluated in terms of mutations in the epitope regions. **Figure 1A** describes the antigenic distances of various strains with respect to X31, for the known epitopes A, B, and C. The antigenic distance was determined in terms of $p_{epitope}$. It can be seen from the graph that antigenic distance is least between the strains X31 and ALB70 with no mutations in the epitopes A and C. Comparatively, the antigenic distance is higher between X31 and PC73 with many mutations in the epitope regions. The antigenic distances of MP98, MOS99, WIS05, BR07 and PT09 with respect to X31 are high for all the epitopes. Maximum antigenic variability is observed in the epitope C within the data set considered (**Figure 1A**). The total number of mutations between the successive strains has been enlisted in **Table 1 (see Supplementary material)**. **Figure 1B** describes the antigenic distances between pairs of NA proteins from successive strains. While only the epitope B changed between the strains X31 and ALB70, all the epitopes differed between ALB70 and PC73. It should be noted that between the strains of MP98 and MOS99, variability was observed in epitopes B and C ($p_{epitope}$ values: 0.19 and 0.48 respectively). No changes were observed in any of the epitopes of NA protein between the strains BR07 and PT09. In order to investigate the molecular basis of the reported failure of Mem5 antibody (which is specific to epitope B of MP98) to bind to NA of strains evolved in 1999 and thereafter **[12]**, we undertook molecular docking studies involving the known structure of Mem5 against predicted 3D structures of NA proteins. Three dimensional structures of the NA proteins from strains MOS99 and WIS05 were predicted using the homology modelling protocol as implemented by SwissModel online workstation (automated mode). The NA protein of influenza (H3N2) strain MP98 was used as a template (PDB ID: 2AEP). Analyses of energy values and Ramachandran plot revealed that the models were reliable **(Table 2, see Supplementary material)**. Docking of Mem5 antibody onto the 3D structures of NA proteins was carried out to determine the binding site and the residues interacting within the antigen-antibody complex through hydrogen bonds (H-bonds) or salt bridges.

The results of docking studies indicated that the antibody Mem5 failed to recognize and bind to the epitope B on the NA protein of MOS99 and strains evolved thereafter. Docking outputs revealed physically unrealistic orientations of antigen with antibody **(Figure S4, see Supplementary material)**. The docking output revealed that Mem5 docked onto NA of MOS99 with the CDR regions interfacing with the amino acid stretch 170-176 on MOS99 but without H-bonds. However, this region on the NA would remain embedded in the tetramer complex on the virion surface and therefore remain inaccessible to antibodies, as observed in the tetrameric structure of NA protein (PDB ID: 2AEP) **(Figure S4B, see Supplementary material)**. The failure of the antibody to recognize epitope B on the NA of MOS99 can be attributed to the variations in amino acid composition (greater antigenic distance). The mutation H197D at the antigenic site B in MOS99 compared to MP98 resulted in alteration of molecular surface contour as well as electrostatics **(Figure 1C)**. This mutation was retained in the subsequent strains WIS05, BR07 and PT09. The amino acid H197 of NA MP98 was crucial for interactions with the Antibody Mem5. In the co-crystal structure of Mem5 with NA of MP98 (2AEP.pdb), the residue H197 on epitope B formed H-bonds with residues R52 and T56 of the heavy chain of Mem5

(data obtained from CSU **[20]** analyses). Hence, the mutation H197D has not only altered the surface electrostatics, but also altered the ability to form H-bonds with the antibody. Another two mutations E199K and K221E at the epitope B were observed in WIS05 and retained thereafter. These mutations along-with another one G216V in the vicinity of epitope B, further altered the side-chain conformation (protein surface) and the surface electrostatics **(Figure 1C)**. From predominantly negatively charged in MOS99, the epitope B has become predominantly positive in WIS05. However, in the absence of suitable antibody structures (1999 onwards) it was impossible to ascertain the impacts of these mutations in terms of antigen-antibody interactions.

Overall, antigenic variability in NA proteins between the various strains of H3N2 viruses have been studied. Though the epitopes of ALB70 were very similar to that on X31, considerable differences existed between those on X31 and other strains evolving from 1973 onwards. Maximum variability was observed in epitope B between the successive evolving strains during 1998-2005. The docking of Mem5, an antibody specific to epitope B of MP98, onto the MOS99 and WIS05 NAs revealed that the antibody fails to recognize the epitope. The loss of recognition by the antibody can be attributed to the mutations on the B epitope of MOS99 and WIS05 which altered the protein surface and the surface electrostatics. A comparison with earlier reports from the authors' laboratory **[11]** revealed that the antigenic variability in NA followed trends similar to that in case of the HA proteins and also indicated the equivalent net divergence within NA proteins. The findings at the molecular level supports the earlier experimental reports related to NA and host immune response **[8, 9]**.

Conclusion:

The antigenic evolution of Neuraminidase proteins from vaccine strains of influenza A/H3N2 has been studied over the period 1968-2009 and variability in terms of antigenic distances have been observed for all the epitopes. The structural basis for the antibody Mem5 not recognizing the NAs of 1999 and subsequently evolved strains could be explained through molecular docking studies. A single amino acid change was found to alter the surface properties, antigenicity of the epitope B in NA proteins and the specificity to antibody binding. The present study may contribute in gaining insight into the mechanisms of virus-host interactions at the molecular level. Our study supports the earlier experimental observations regarding NA versus host-immune response and reinforces the suggestions that the NA protein may be considered along with HA as an active component for effective influenza vaccine design.

References:

[1] Muñoz ET & Deem MW. *Vaccine* 2005 **23**: 1144 [PMID: 15629357]
[2] Bush RM *et al*. *Science* 1999 **286**: 1921 [PMID: 10583948]
[3] Bush RM *et al*. *Mol Biol Evol*. 1999 **16**: 1457 [PMID: 10555276]
[4] Lee MS & Chen JS. *Emerg Infect Dis*. 2004 **10**: 1385 [PMID: 15496238]
[5] Smith DJ *et al*. *Science* 2004 **305**: 371 [PMID: 15218094]
[6] Plotkin JB & Dushoff J. *Proc Natl Acad Sci U S A*. 2003 **100**: 7152 [PMID: 12748378]
[7] Colman PM. *Protein Sci*. 1994 **3**: 1687 [PMID: 7849585]
[8] Webster RG *et al*. *Virology* 1988 **164**: 230 [PMID: 2452514]
[9] Johansson BE *et al*. *Vaccine* 1998 **16**: 1009 [PMID: 9682352]
[10] Dolan S *et al*. *Center Dis Control Prevent Morb Mortal Wkly Rep*. 2004 **53**: 8
[11] Cherian S *et al*. *Bioinformation* 2011 **6**: 35 [PMID: 21464843]
[12] Gulati U *et al*. *J Virol*. 2002 **76**: 12274 [PMID: 12414967]
[13] Tamura K *et al*. *Mol Biol Evol*. 2007 **24**: 1596 [PMID: 17488738]
[14] Shil P *et al*. *In Silico Biol*. 2006 **6**: 373 [PMID: 17274766]
[15] Kolaskar AS & Tongaonkar PC. *FEBS Lett*. 1990 **276**: 172 [PMID: 1702393]
[16] http://spdbv.vital-it.ch/
[17] http://www.ebi.ac.uk/thornton-srv/software/PROCHECK
[18] http://noch.sourceforge.net/
[19] http://zdock.bu.edu/
[20] http://bip.weizmann.ac.il/oca-bin/lpccsu

Edited by P Kangueane
Citation: Shil *et al.* Bioinformation 7(2): 76-81 (2011)

BIOINFORMATION

open access

Supplementary material:

Table 1: List of mutations in NA protein between successive strains

Strain →	X31 vs ALB70	ALB70 vs PC73	MP98 vs MOS99	MOS99 vs WIS05	WIS05 vs BR07	BR07 vs PT09
Total number of mutations	16	13	9	13	9	3

Table 2: Analyses of 3D structures of NA proteins

Strain Name	Sequence identity with template(%)	Minimized Energy (kJ/mol)	RMSD (Å) with template	Occupancy* (%) in Ramachandran plot
MOS99	97.68	-21874.86	0.06	100
WIS05	95.62	-22159.46	0.09	100
2AEP.pdb (Template)	-	-21561.28	-	100

*N.B. Occupancy in Ramachandran Plot denotes the percentage of amino acids in the favourable and generously allowed regions.

	X31	ALB70	PC73	MP98	MOS99	WIS05	BR07	PT09
X31	100							
ALB70	96.6	100						
PC73	94.2	97.2	100					
MP98	88.7	91.3	92.8	100				
MOS99	88.1	90.4	91.5	97.9	100			
WIS05	86.6	88.7	89.8	95.3	97.2	100		
BR07	85.7	87.4	88.9	94.2	95.7	98.1	100	
PT09	85.5	87.2	88.7	94.0	95.5	97.9	99.4	100

Figure S1: Identity (%) between Influenza A/H3N2 NA amino acid sequences. Table generated as output of ISHAN package.

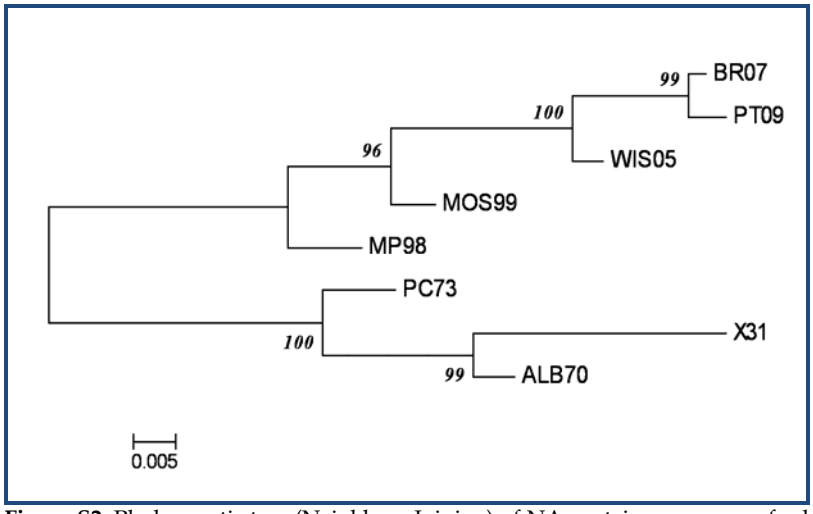

Figure S2: Phylogenetic tree (Neighbor –Joining) of NA protein sequences of selected H3N2 strains (with bootstrap analysis as test of phylogeny with 10000 replications) as obtained from MEGA 4.0.

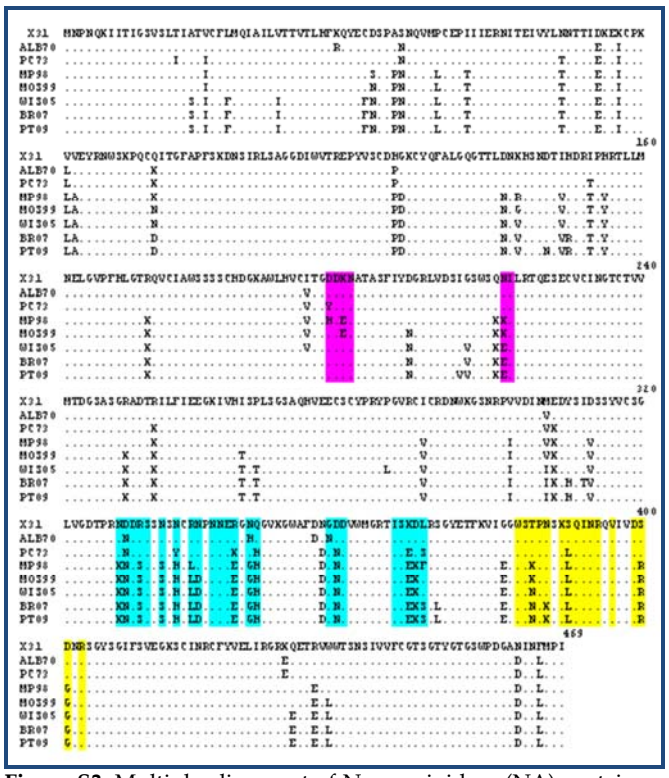

Figure S3: Multiple alignment of Neuraminidase (NA) protein sequences from H3N2 strains. Epitopes A, B and C are highlighted in Yellow, Pink and Cyan respectively.

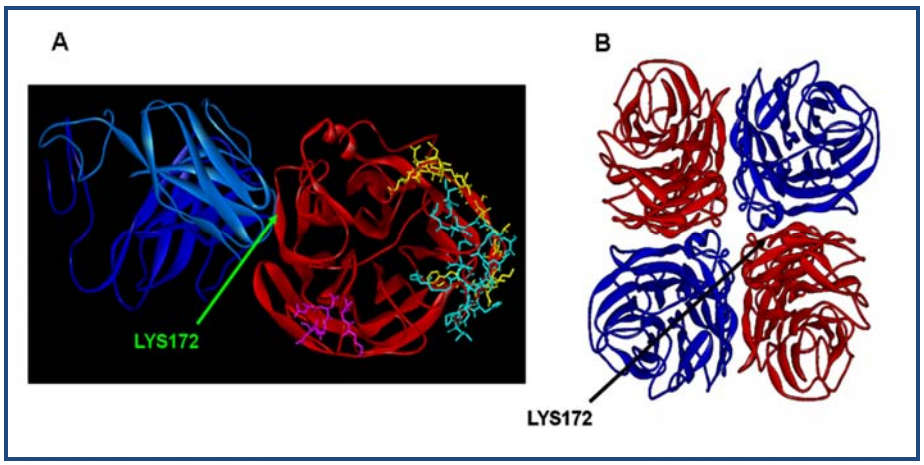

Figure S4: Physically unrealistic orientation of Antibody Mem5 (Blue) docked with NA Antigen (red) with epitope regions highlighted A (yellow), B (Pink) and C (Cyan) residues at the docking interface shown in (A) remain inaccessible in the natural tetramer of the NA protein (B).