ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 5, Issue, 08, pp.2135-2139, August, 2013

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

A PROTEOMIC ANALYSIS AND STRUCTURE VALIDATION OF HEMAGGLUTININ AND NEURAMINIDASE GENE AMONG THE EPIDEMIC STRAINS OF H1N1

Priya Roosvelt, Indu Purushothaman and *Rajarajan, S.

Bioinformatics Infrastructure Facility Centre of DBT (BIFC), Presidency College (Aut), Chennai, India

ARTICLE INFO

Article History:
Received 24th May, 2013
Received in revised form
02th June, 2013
Accepted 18th July, 2013
Published online 23rd August, 2013

Key words:

Hemagglutinin, Neuraminidase,
Antigenicity, Automated modeling.

ABSTRACT

Since its first origin in 1918, Influenza virus has caused the highest degree of morbidity and mortality compared with the other dreadful infections. Till date, many works were carried out on the molecular level to unveil the mystery behind this dreadful infection. In this paper, we have carried out a proteomic and structure validation analysis for the two important genes, Hemagglutinin and Neuraminidase of influenza virus isolated between 1918 to 2009. Among the subjected isolates, HA and NA protein of 2009 isolates were highly antigenic. The subcellular location of all pretense was located in the plasma membrane. Simultaneously, the structure quality checks inferred that Phyre 2 produced a better and a significant prediction for ten isolates of Hemagglutinin and Neuraminidase protein when compared to Raptor X and (PS) ^{2} software.

Copyright © 2013 Priya Roosvelt, *et al.*, This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

It has been well established that epidemics have been caused by influenza every year. The recent one is the outburst of H7N9 in China. Based on the type of Hemagglutinin (HA) and neuraminidase (NA) present in the genome of Influenza A virus, the subtype causing the infection is determined. In 1990, Wiley and his co-workers (Wiley *et al.*, 1990) obtained the purified X-ray structure of hemagglutinin by working on the original structure obtained by Wilson in 1981 (Wilson *et al.*, 1981). The main functions of Hemagglutinin are: helps the viruses in its attachment and diffusion into the host cells. The host neutralization antibody reaction is motivated by this antigenic protein. Till date, seventeen different HA antigens have been identified. H1NI influenza viruses are referred to as swine flu. It appears as a non-covalent homo-timer known as HA0 when it is translocated transversely around the rough endoplasmic reticulum membrane (Braakman *et al.*, 1991) and is implicated by RNA segment 4 (Palese *et al.*, 1976). HA0 is dissected by cellular proteases to form c-terminus three hundred twenty seven amino acid residues (HA1) and N-terminus two hundred and twenty two amino acid residue (HA2) after it moves to the plasma membrane via the Golgi complex (Copeland *et al.*, 1986). HA1 and HA2 are joined by two intramonomer disulphide bonds. With the help of a coiled transmembrane peptide (27 amino acids) which is present close to c-terminus of HA2 chain, each monomer is attached in the viral membrane (Weis *et al.*, 1988). The progeny virus release and spread of the infection to neighboring cells from the infected cell surface is assisted by the neuraminidase helping in removing the sialic acid during the final stage of infection of influenza virus (Bucher and Palese, 1975). Two out of nine NA (N1 and N2) subtypes belonging to the avian cause's deadly disease in humans. Till date, the 1918 "Spanish" H1N1 was considered to be dreadful which brought a mortality of twenty to fifty million people worldwide (Crosby, 1989). The mutation was found at high frequency in HA and NA genes that resulted in regular changes in the antigen present on the surface proteins (Fitch *et al.*, 2000). New mutations positions were identified after careful analysis of 23 Indian A/H1N1 2009 viral HA gene and found to be similar to A/California/07/2009 (Gunasekeran *et al*., 2012).

***Corresponding author:** Rajarajan, S.
Bioinformatics Infrastructure Facility Centre of DBT (BIFC), Presidency College (Aut), Chennai, India

An atomic level structural study on Neuraminidase was performed to understand the preservation of a D147-H/R150 salt bridge which is essential for the constancy of 150 cavities present in the active site of group 1 NA and thus help in designing antiviral compounds to act on the 150 cavities resulting in the inactivation of the 2009 H1N1 pandemic and avian H5N1 viruses (Amaro *et al.*, 2011). Linear epitope (P5) was identified with the H1N1 pdm virus by peptide scanning. This epitope was found to be immunodominant, highly conserved and not present in other subtypes and can be used for sero-diagnosis of the disease (Zhao *et al.*, 2011). In the present study, we have retrieved the HA and NA sequence isolated from 1918 to 2009 in NCBI to understand its antigencity, stability and subcellular location followed by sequence analysis. Further, homology modeling was performed using Phyre v2 and cross verified using latest bioinformaitcs softwares followed by its quality checked.

MATERIALS AND METHODS

Sequence retrieval

The HA gene sequences submitted during 1918 till 2009 in NCBI with accession no AAD17229, BV25634, AAA43171, ACD88516, ABF21276, AAP34323, AAK70456, AFP33324, ABY40408, AFF57188 and NA gene sequences with accession no: AAF77036, AAF77044, AAA43797, ACD88519, ABF21332, AFR76584, AA088264, AFR76062, ABY40396 and ADZ53141 were retrieved from NCBI.

Sequence analysis

Multiple sequence alignment was performed among the sequences of HA and gene and NA gene for the Visual depictions of the alignment to understand the mutational events such as point mutations and insertion or deletion mutations in the sequence. Multiple sequence alignment is often used to assess sequence conservation using Praline. The effect of these converted individual amino acids on the stability of the whole protein was studied by Align GVGD tool (http://agvgd.iarc.fr/).

Prediction of its antigenicity, subcellular localization and signal peptide

The antigenicity of each protein was calculated using VaxiJen v2. 0 at a threshold of 0.5%. The subcellular location of the protein was predicted using Virus- pLoc. The presence of signal peptide was identified using SignalP v 3.0.

Homology Modeling

Phyre2

Phyre and Phyre2 (Protein Homology/AnalogY Recognition Engine) are web-based services for automated homology modelling.

RaptorX

A protein structure prediction server developed by Xu group, excelling at predicting 3D structures for protein sequences without close homologs in the Protein Data Bank (PDB). For the input sequence it predicts its secondary and tertiary structures as well as solvent accessibility and disordered regions

PS^2 (Protein Structure Prediction Server)

An automated homology modeling server. It uses an effective consensus strategy by combining PSI-BLAST, IMPALA, and T-Coffee in both template selection and target-template alignment. The final three dimensional structure is built using the modeling package MODELLER.

Structure Validation by Errat

Structure validation was performed using SAVes version 4 (Structural analysis and verification server) modules like Errat was used to evaluate the quality of the structure. Errat defines the overall quality factor of a protein.

ProQ

ProQ (http://www.sbc.su.se/~bjornw/ProQ/ProQ.html) is a neutral network based predictor that based on a number of structural features predicts the quality of a protein model. Is optimized to find correct models based on the LG score and MaxSub. LG score is -log of a P-value and MaxSub ranges from 0-1, where 0 is insignificant and 1 very significant.

RESULTS

The HA gene isolated from Korea in 2009 was found to be highly antigenic (0.5507) when compared with the rest of the other HA genes. The sub-cellular location of the predicted HA was found to be in the plasma membrane. Signal peptide was found in all the HA genes isolated during the period from 1918 till 2009 (Table 1). While in the case of NA gene, isolated from Hong Kong in 2009 was found to be highly antigenic (0.5364) compared with 1918 gene which was found to be non-antigenic. The sub –cellular location of NA protein was also in plasma membrane same as of HA genes. Signal peptide was present in 1918, 1930,1954,1995,1998 and 2000 (Table 2). The comparative analysis performed among the HA and NA gene isolated during 1918 till 2009 revealed that there was a notable difference in the protein isolate of 1918 first pandemic compared with the 2009 isolate. The position at which the amino acid conversion had occurred were noted (Table 3 and 4) and its significance in the aspect of protein stability was checked using Align GVGD software. The overall protein structure of Hemagglutinin and neuraminidase developed by Phyre2, Raptor X and PS2 was checked for its quality factor by Errat. It was found that 1991(87.62%) and 1995 (60.94%) hemagglutinin isolates developed by Phyre 2 had the highest and lowest scoring rate respectively. While, 69.92% (2000) and 55.62% (1980) was predicted by Raptor X and PS2 result showed that 79.23% was seen in 1930 and 69.75% was scored by 1954 isolates.

Table 1. Prediction of the overall quality factor, antigenic value, location and presence of signal peptide Hemagglutinin gene of Influenza virus H1N1

Year of isolation	Place of isolation	Accession number	Antigenic value	location	Prediction of Signal peptide
1918	South Carolina	AAD17229	0.5464	Plasma membrane	yes
1930	Iowa	ABV25634	0.5514	Plasma membrane	yes
1954	Leningrad	AAA43171	0.5333	Plasma membrane	yes
1980	Kansas	ACD88516	0.5309	Plasma membrane	yes
1991	Texas	ABF21276	0.5085	Plasma membrane	yes
1995	Bejing	AAP34323	0.5247	Plasma membrane	yes
1998	HongKong	AAK70456	0.5115	Plasma membrane	yes
2000	England	AFP33324	0.5006	Plasma membrane	yes
2005	Chonburi	ABY40408	0.5195	Plasma membrane	yes
2009	Korea	AFF57188	0.5507	Plasma membrane	yes

Table 2. Prediction of the overall quality factor, antigenic value, location and presence of signal peptide of Neuraminidase gene of Influenza virus H1N1

Year of isolation	Place of isolation	Accession number	Antigenic value	location	Prediction of Signal peptide
1918	Bervig Mission	AAF77036	0.4801	Plasma membrane	yes
1930	Iowa	AAF77044	0.4733	Plasma membrane	Yes
1954	Leningrad	AAA43797	0.5009	Plasma membrane	Yes
1980	Kansas	ACD88519	0.5008	Plasma membrane	Yes
1991	Texas	ABF21332	0.4921	Plasma membrane	No
1995	Bejing	AFR76584	0.5167	Plasma membrane	yes
1998	HongKong	AAO88264	0.5007	Plasma membrane	yes
2000	England	AFR76062	0.5106	Plasma membrane	yes
2005	Chonburi	ABY40396	0.5379	Plasma membrane	No
2009	HongKong	ADZ53141	0.5364	Plasma membrane	No

Table 3. Comparison of amino acid sequence of HA belonging to different strain of H1N1 isolates

	Position of substitution at the following amino acid residues at HA																						
Year of Isolation	4	16	53	60	62	64	68	86	88	90	111	137	144	146	156	172	187	202	203	211	219	221	232
1918	R	N	S	K	K	I	Q	L	L	A	D	T	E	T	A	G	G	G	T	N	G	K	A
1930	I	N	S	R	G	I	Q	L	L	V	D	T	E	T	A	E	G	S	T	N	G	K	A
1954	K	D	S	R	K	I	Q	S	V	K	Y	E	N	T	K	N	E	I	E	K	V	N	E
1980	I	N	R	K	G	I	H	L	F	V	D	A	E	N	E	E	G	S	T	N	G	K	A
1991	K	Y	S	R	K	I	Q	S	F	K	Y	E	T	T	E	N	E	I	G	T	V	H	K
1995	K	Y	S	L	K	I	Q	S	I	K	Y	E	T	-	E	N	E	I	G	T	V	H	K
1998	K	Y	S	L	K	I	Q	L	I	K	Y	E	T	-	K	N	E	I	G	T	V	H	K
2000	K	Y	S	R	K	T	Q	S	F	K	Y	D	T	T	K	N	G	M	G	K	L	H	K
2005	I	N	R	N	R	K	H	L	F	V	D	A	E	N	T	G	K	N	T	N	G	K	E
2009	I	N	K	K	R	V	H	S	S	A	D	T	D	N	A	G	G	S	D	N	G	R	I

Table 4. Comparison of amino acid sequence of NA belonging to different strain of H1N1 isolates

	Position of substitution at the following amino acid residues at NA												
Year of isoaltion	16	20	46	79	84	189	222	264	287	289	332	365	435
1918	V	I	P	D	I	G	N	T	K	M	T	T	-
1930	I	T	P	D	I	S	K	I	K	M	T	I	-
1954	A	I	T	D	T	G	Q	T	T	M	K	N	T
1980	I	V	P	G	I	G	K	I	K	V	K	I	-
1991	A	I	T	D	T	G	R	T	T	M	E	N	T
1995	T	A	P	A	K	S	N	V	E	T	T	T	-
1998	I	V	S	G	I	G	K	I	K	V	R	I	-
2000	T	A	P	T	R	N	N	V	V	T	T	T	-
2005	T	A	P	A	K	S	N	V	K	T	T	T	-
2009	T	A	I	S	K	N	N	V	E	T	T	I	-

Table 5. Validation of Hemagglutinin and Neuraminidase protein by Errat

Year of Isolation	Hemagglutinin protein			Neuraminidase protein		
	PHYRE2	RAPTOR X	PS2	PHYRE2	RAPTOR X	PS2
1918	82.98	60.54	77.32	85.68	86.12	77.18
1930	81.03	57.78	79.23	84.96	80.89	78.15
1954	79.30	63.20	69.75	80.00	78.59	71.69
1980	86.11	55.62	75.24	85.75	76.44	74.54
1991	86.62	60.23	76.19	80.26	78.59	78.57
1995	60.94	65.10	73.80	87.00	81.77	80.64
1998	84.65	58.63	71.97	84.43	75.39	74.80
2000	70.78	69.92	71.19	86.47	77.75	76.13
2005	85.89	58.36	76.82	87.80	80.47	75.06
2009	84.14	58.07	73.48	86.44	82.24	75.86

Table 6. Hemagglutinin and Neuraminidase protein structure quality check by ProQ (Protein quality predictor)

Year of isolation	Hemagglutinin protein structure quality			Neuraminidase protein structure quality		
	LG score			LG score		
	Phyre	Raptor X	$(PS)^2$	Phyre	Raptor X	$(PS)^2$
1918	4.37	3.12	3.95	4.66	4.65	4.95
1930	4.28	3.23	3.84	4.56	4.79	4.70
1954	3.96	3.38	4.20	4.45	4.79	4.38
1980	4.12	3.35	4.27	4.97	4.34	4.58
1991	4.03	3.32	4.13	4.59	4.27	4.72
1995	3.82	2.91	3.99	4.63	4.73	4.82
1998	4.12	3.45	4.23	4.60	4.44	4.90
2000	3.92	3.33	4.01	4.60	4.79	4.76
2005	4.27	3.25	4.23	4.70	4.78	4.40
2009	4.05	3.19	3.99	4.72	4.48	4.72

Different ranges of quality: LGscore>1.5 fairly good model, LGscore>2.5 very good model, LGscore>4 extremly good model

In the case of neuraminidase protein, 2005 isolates scored the highest 87.8% while the 1954 isolates secured 80% by Phyre 2 modeling software. As per Raptor X, 1918 isolates produced highest score (86.12%) and 1980 predicting the least score (76.44%) and in the case of PS2 tool, the 1995(80.64%) and 1954 (71.69%) isolates had the highest and least quality protein predicted based on the quality check carried out by Errat (Table 5). The protein structure of Hemagglutinin and neuraminidase developed by Phyre2, Raptor X and PS2 was checked for its quality by ProQ. It was found that 1918 (4.37) and 2000 (3.92) hemagglutinin isolates developed by Phyre 2 had the highest and lowest LG score respectively. While, 1998 (3.45) and 1995 (2.91) isolates showing the highest and least LG score predicted by Raptor X and PS^2 result showed that 4.27 (1980) and 3.95 (1918) sharing the highest and lowest LG score. In the case of neuraminidase protein, 1980 isolates (4.97) and the 1954 isolates (4.45) were the highest and lowest quality protein predicted by Phyre 2 modeling software. Raptor X was able to produce 1930, 1954, 2000 isolates of top score (4.79) and 1991 with the least score (4.27) and $(PS)^2$ tool, the 1954 isolates (5.15) and 2005 (4.40) was the highest and lowest quality protein predicted. (Table 6).

DISCUSSION

A comparative analysis among the HA genes was performed and from the result, it was inferred that a significant variation was seen in the protein isolated in 1918, the first pandemic strain of influenza virus when compared with the 2009 isolate. Based on the result obtained, significant variation was found between the HA protein isolated from 1918 to 2009 which resulted in the increase of virulence yet the stability of the proteins was not affected. The three dimensional structure of the epidemic strains were modeled using Phyre2 and was cross verified with Raptor and PS^2. The overall quality of the refined protein structure was checked using Errat (Figure1 and 2) and ProQ (Figure 3 and 4) which showed that the structure generated using Phyre2 was of good quality. The highest value (86.6%) 1991 HA gene and the lowest value was (60.9%) 1995. It was noted the changes in the amino acid does not bring in an big impact on the stability of the protein. Till date, no studies had been carried out on quality check of Hemagglutinin and neuraminidase protein structure and in identification of virulent strains. This study has gained its importance as previous studies in the recent decades was based on the effectively control of present pandemic H1N1 virus 2009 by use of Oseltamivir and Zanamivir drug which was inferred by the *in silico* analysis of different drugs tested against the swine flu virus, compared with 1918 H1N1 viral proteins (Singh *et al*., 2009).

The Influenza research database (IRD) was designed by the U.S National Institute of Allergy and Infectious diseases to enlightens with information regarding epitope, genomic and proteomic data, data regarding host range restriction and transmission of 2009 pandemic H1N1 influenza virus to enable scientists to develop an effective treatment towards influenza infection (Squires *et al.*, 2012). The molecular modeling for novel receptors using Schrodinger and docking of suitable inhibitors in influenza proteins was studied using I-Tasser and Modeler 9.9 tool. The study was concluded by developing an eastern Indian homology model of receptors for HA and NA from H1N1 and docking result can be used for new drug design (Behera *et al.*, 2012). A phylogenetic study showed that European H1N1 swine flu was closely related to H5N1 avian flu than H1N1. The study also exposed a new mutation in the protein which is the binding cavity for the currently used NA inhibitors (Stroh *et al.*, 2009).

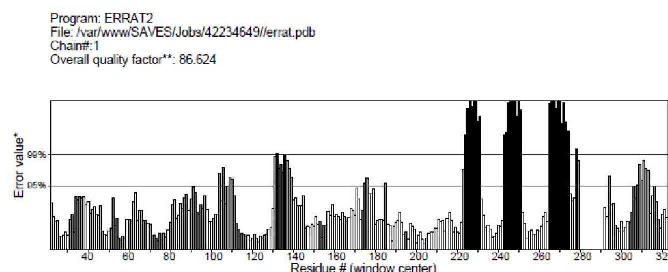

Figure 1. Errat score for Hemaggluttinin protein of 1991 isolates

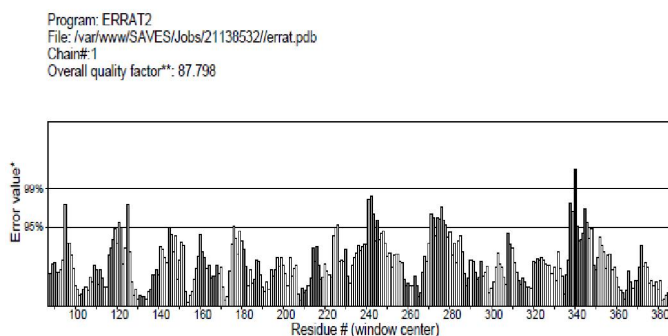

Figure 2. Errat score for Neuraminiasde protein of 2005 isolates

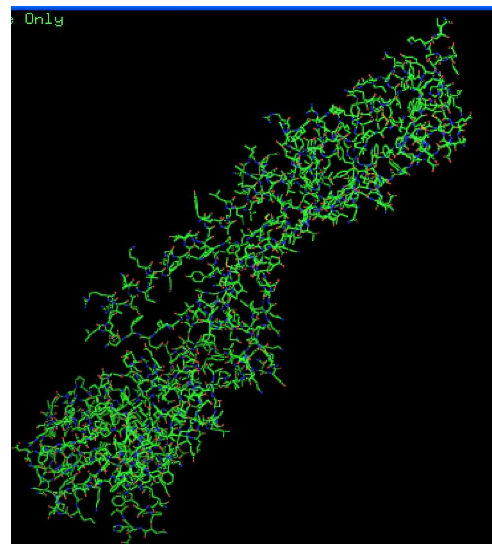

Figure 3. Pro Q quality checked for Hemaggluttinin protein of 1918 isolates

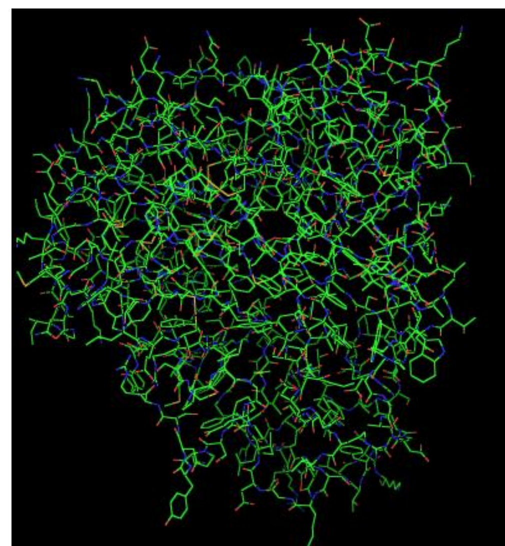

Figure 4. Pro Q quality checked for Hemaggluttinin protein of 1980 isolates

Conclusion

From this study, the virulent strain for both genes were identified which can be used for further analysis. It was concluded that the prediction of 3D structure of ten HA and NA proteins will help in better understanding the mutational changes undergone from the first isolates till pandemic one. It is also useful in unveiling the mystery behinds the molecular function of each protein.

REFERENCE

Amaro, RE. Swift, RV. Votapka, L. Li, WW. Walker, RC. and Bush, RM. 2011. Mechanism of 150-cavity formation in influenza neuraminidase Nature Communications., 2: 388.

Behera, DK. Behera, PM. Acharya, L. Dixit, A. and Padhi P. 2012. *In Silico* biology of H1N1: Molecular modelling of novel receptors and Docking Studies of inhibitors to reveal new insight in Flu treatment. Journal of Biomedicine and Biotechnology, 2012:1-7.

Braakman, I. Hoover-Litty, H. Wagner, KR. and Helenius, A. 1991. Folding of influenza hemagglutinin in the endoplasmic reticulum. J. Cell Biol., 114: 401–411.

Bucher, D, and Palese, P. 1975. The biologically active proteins of influenza virus: neuraminidase, p. 83–123. *In* E. D. Kilbourne (ed.), The influenza viruses and influenza. Academic Press, New York, NY.

Copeland, CS, Doms, RW, Bolzau, EM, Webster, RG. and Helenius, A. 1986. Assembly of influenza hemagglutinin trimers and its role in intracellular transport. J. Cell Biol. 103: 1179– 1191.

Crosby, AW. 1989. America's forgotten pandemic: the influenza of 1918. Cambridge University Press, Cambridge, United Kingdom.

Fitch, WM. Bush, RM. Bender, CA. Subbarao, K. Cox, NJ. 2000. Predicting the evolution of human influenza A. J Hered.,91:183-5.

Gunasekeran, P. Krishnaswamy, K. Arunagiri, K. Sambasivam, M. Lakshmipathy, M. Arunpon, Fatima, SG. 2012. Genetic analysis of HA gene of pandemic H1N1 2009 influenza viruses circulating in India. Indian Journal of Medical Microbiology., 30(3):346-9.

Palese, P. and Schulman, JL. 1976. Mapping of the influenza virus genome: identification of the hemagglutinin and the neuraminidase genes. Proc. Natl. Acad.Sci.U.S.A.,73:2142–2146.

Singh, KhD. and Muthusamy, K. 2009. *In silico* analysis and drug efficacy test of influenza A virus (H1N1) 2009. Indian. J. Microbiol., 49:358-364.

Squires, RB. Noronha, J. Hunt, V. Garcı´a-Sastre, A. Macken, C. Baumgarth, N. Suarez, D. Pickett, BE. Zhang, Y. Larsen, CN. Ramsey, A. Zhou, L. Zaremba, S. Kumar, S. Deitrich, J. Klem, E. Scheuermann, RH. 2012. Influenza Research Database: an integrated bioinformatics resource for influenza research and surveillance. Influenza and Other Respiratory Viruses, 9(6): 404-416.

Stroh, SM. Ma, J. Lee, RTC. Sirota, FL and Eisenhaber, F.2009. Mapping the sequence mutations of the 2009 H1N1 influenza A virus neuraminidase relative to drug and antibody binding sites. Biology Direct., 4: 9.

Weis, W. Brown, JH. Cusack, SC. Paulson, JC. Skehel, J.J. And Wiley, DC. 1988. Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. Nature., 333 : 426–431.

Weis, WI. Brünger, AT. Skehel, J.J. and Wiley, DC. 1990. Refinement of the influenza virus hemagglutinin by simulated annealing. J. Mol. Biol., 212: 737–761.

Wilson, I.A. Skehel, JJ. Wiley, DC. 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature., 289: 366–373.

Zhao, R. Cui, S. Guo, L. Wu, C. Gonzalez, R. Paranhos-Baccala, G. Vernet, G. Wang, J. Hung, T. 2001. Identification of a highly conserved H1 subtype-specific epitope with diagnostic potential in the Hemagglutinin protein of Influenza A Virus. PLoS One.,6 (8):e23374.

****