Innovare Academic Sciences

International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 7, Issue 4, 2015

Original Article

A GLOBAL COMPARISON AND ANALYSIS OF NEURAMINIDASE H1N1 STRAIN OF INFLUENZA A

PRASHANT SAXENA[1], SHASHANK SHRISHRIMAL[2], INAMUL HASAN MADAR[3]

[1]Department of Bioinformatics, Sathyabama University, Chennai, India, [2]Department of Biochemistry and Molecular Biology, University of Nebraska Medical Center, Omaha, NE 68198-5870, USA, [3]Bioinformatics Bishop Heber College, Tiruchirappalli, Tamil Nadu, India. Email: shashank.shrishrimal@unmc.edu

Received: 27 Nov 2014 Revised and Accepted: 25 Dec 2014

ABSTRACT

Objective: To evaluate the variation of neuraminidase (NA) protein from various H1N1 strain for drug designing.

Methods: In this study we have used 12 sequences of NA protein from various countries, retrieved from Uniprot KB database. We have performed structural analysis, antigenic and glycosylation site prediction between NA proteins of influenza A strains.

Results: Antigenic variants in sequence of NA H1N1 strain from Italy were found to be unique and were not present in any other NA H1N1. Strains from Italy and Thailand were found to be distantly related while others are closely related. We observed the maximum similarity from position 84 to 448 using disorder prediction analysis of different strains. Sequences from 1 to 83 and 449 to 469 showed the maximum dissimilarity among the NA Proteins.

Conclusion: This study focuses on the regions of sequence similarity and dissimilarity of NA in H1N1 strains from different countries. The results in this paper are based on currently available sequences for NA of H1N1 strains and bioinformatic tools. Our study will help in understanding of the regions of high variability due to mutations and conserved domains that can be potential targets for drug development.

Keywords: Influenza, Neuraminidase, Glycosylation, Vaccine.

INTRODUCTION

H1N1 has caused widespread outbreaks, including epidemics and pandemics, of acute upper or lower respiratory tract infection. The 2009 re-emergence of the strain led to declaration of a global H1N1 pandemic by the World Health Organization (WHO) and was the first ever, global pandemic since the 1968 Hong Kong flu. By 2010, the strain had spread to more than 214 countries and caused 18, 138 deaths [1]. This is strain popularly known as "Swine Flu" (swine influenza) because of its origins from pigs by genetic re-assortment [2].

H1N1 also called influenza type A subtype H1N1, belongs to the family of Orthomyxoviridae. The segmented RNA genome of the virus consists of eight strands, made from a single species or multiple species that confers the virus cross-species infectivity. Influenza subtype type A H1N1 strains consist of one strand derived from human flu strains, two from avian (bird) strains, and five from swine strains. The nucleic acid of influenza virus translated into approximately 10 proteins, out of which two are viral membrane glycoproteins: hemagglutinin (H) and neuraminidase (N). They are used to classify the different subtypes of Influenza A virus [1]. They are essential for viral infection and release from infected cells. There are 16 known H proteins and 9 known N proteins, forming different subtypes like the H1N1 subtype. The name H1N1 corresponds to hemagglutinin type 1 (H1) and neuraminidase type 1 (N1) antigens, present on the viral coat [3].

Neuraminidase a major surface glycoprotein, possess enzymatic activity that cleaves the Sialic acid receptor, enabling the virus to release from the host cell after replication. Most of the anti-influenza drugs target the neuraminidase (NA) protein, inhibiting its function and providing some treatment. However, drug resistant mutant strains emerge, limiting the capacity of treatment by most drugs. Mutant strains are also resistant to antibodies generated from the available vaccines [4-7]. The need to design new influenza vaccines every year to keep up with the new strains is challenging and an understanding of the variations in the strains is critical for designing future drugs.

In our study, we evaluate the variation of the neuraminidase protein from various H1N1 strain by comparative analysis. Nucleotide and protein sequence similarity analysis performed using the available tools such as T-coffee, Garnier, Pepstats. We also perform analysis of glycosylation and antigenic variants among different protein sequences of NA H1N1 strain. Divergence among the sequences of NA receptor protein ofH1N1has been done by construction of the phylogenetic tree, using the distance method. Protein disorder analysis helps in understanding the function of short motifs.

MATERIALS AND METHODS

Collection of influenza NA H1N1 strain sequence

The protein sequences of NA H1N1 strain among various countries retrieved [8] from Uniprot KB database [9]. The sequences are from the following countries: Malaysia (1954), New Jersey(USA)(1976), India (1980), Memphis (USA)(1996), New Zealand (2000), Russia (2006), Italy (2009), Thailand (2010), Texas(USA)(2010), Brazil (2011), China (2012) and Kenya (2013).

Sequence similarity analysis

Multiple sequence alignment (MSA) technique [10] was used to identify the divergence and mutations in the protein sequence of NA H1N1 strain obtained from various countries. T-Coffee tool [11] (http: //tcoffee. crg. cat/apps/tcoffee/do: expresso) was used to obtain the MSA. It generates the multiple sequence alignment on the basis of pair wise alignment between possible pairs of the sequence.

Protein sequence analysis

Primary and secondary structure analysis of all the protein sequences was done using Garnier and Pepstats tool from EMBOSS 2.10.0-0.8 package [12].

Analysis of glycosylation and antigenic variants

Variations in the NA glycosylation sites were determined by using NetNGlyc [13] 1.0 online tool. The antigenic divergence was also determined by using CTL Pred tool [14].

Phylogenetic analysis

For the construction of a phylogenetic tree distance method *i.e.* UPGMA (Un weighted Paired Group Method of Arithmetic mean) by MEGA (Molecular Evolutionary Genetics Analysis) 5.2 was used [15].

Disorder analysis

The disorder regions present within the proteins sequences are predicted using the PONDR®s VLXT software [16-18].

RESULTS AND DISCUSSION

Protein sequence analysis

Primary structures of the proteins were analysed by using Pepstats tool from the EMBOSS package. The result indicates (table 1) the maximum similarity among the amino acid composition from various protein sequences of NA H1N1 strain. The secondary structure of protein was analysed by using Garnier tool from the EMBOSS package (table-2) shows, there is a slight variation among the protein sequences.

Table 1: Amino acid composition among various NA H1N1 strains

Amino Acid %	Malaysia	New jersey (USA)	India	Memphis (USA)	New Zealand	Russia	Italy	Thailand	Texas (USA)	Brazil	China	Kenya
Ala	3.4	3.8	3.8	4.3	4.3	4.3	3.8	3.3	3.4	3.4	3.4	3.4
Cys	4.0	4.1	4.0	3.8	3.8	3.8	4.5	4.2	4.1	4.1	4.1	4.1
Asp	6.3	4.6	5.7	5.5	5.1	5.1	4.8	4.0	4.3	4.1	4.3	4.3
Glu	3.6	3.6	3.6	3.8	3.8	3.6	4.5	4.5	4.3	4.3	4.3	4.3
Phe	3.4	3.6	3.4	3.4	3.6	3.6	4.2	4.0	3.8	3.7	3.8	3.8
Gly	9.6	9.8	9.6	9.6	9.6	9.6	10.4	9.6	9.6	9.7	9.6	9.6
His	1.7	1.7	1.7	1.7	1.9	1.9	1.3	1.4	1.3	1.3	1.3	1.3
Ile	9.6	9.8	8.9	8.9	8.9	9.1	8.1	9.4	9.8	9.7	10.0	9.4
Lys	4.5	4.2	4.5	4.2	5.1	5.3	4.3	4.0	4.3	4.1	4.3	4.3
Leu	4.0	5.1	4.0	4.7	4.5	4.5	3.8	4.0	3.8	3.9	3.8	3.8
Met	2.3	2.1	2.1	2.1	1.5	1.7	1.0	1.0	1.5	1.5	1.5	1.5
Asn	6.4	7.0	6.6	7.0	7.4	7.4	8.1	9.1	9.1	9.1	8.3	8.3
Pro	4.7	4.4	4.7	4.5	4.3	4.3	5.1	4.5	4.7	4.7	4.7	4.7
Gln	2.8	2.9	2.6	2.6	2.3	2.6	2.0	3.3	3.2	3.2	3.2	3.2
Arg	4.5	4.0	4.7	4.7	3.6	3.4	4.3	4.0	3.6	3.7	3.6	3.8
Ser	10.4	11.3	10.6	10.2	10.6	10.6	12.4	12.2	11.5	11.4	11.9	12.2
Thr	6.6	6.2	7.0	7.0	7.0	7.0	5.1	5.4	5.5	5.6	5.8	5.8
Val	5.7	4.9	6.0	5.3	6.1	5.7	6.1	6.1	6.1	6.0	5.8	6.0
Trp	3.4	3.4	3.4	3.4	3.4	3.4	3.3	3.3	3.4	3.4	3.4	3.4
Tyr	3.0	3.0	3.0	3.2	3.0	3.0	3.0	3.3	3.0	3.0	3.0	3.0

Table 2: Secondary structure of selected NA H1N1 strains

Viral Strain	Helix (%)	Strand (%)	Turns (%)	Random coil (%)
Malaysia	8.5	33.7	38.7	19.1
New Jersey (USA)	6.5	33.7	37.2	22.6
India	8.1	34.5	37.7	19.8
Memphis (USA)	8.1	32.6	38.9	20.4
New Zealand	10.0	32.6	38.1	20.6
Russia	9.1	32.1	37.4	21.3
Italy	5.8	29.8	38.6	25.8
Thailand	4.7	33.3	37.2	24.8
Texas (USA)	6.6	33.3	35.4	24.7
Brazil	5.4	33.8	36.0	24.8
China	6.6	33.3	35.6	24.5
Kenya	6.6	32.2	37.1	24.1

Table 3: Glycosylation sites in NA H1N1 strains

Malaysia		New jersey (USA)		India		Memphis (USA)	
Position	**Sequence**	**Position**	**Sequence**	**Position**	**Sequence**	**Position**	**Sequence**
63	NQTY	50	NQSV	88	NSSL	88	NSSL
88	NSSL	63	NQTY	146	NGTV	146	NGTV
146	NGTV	68	NISN	235	NGSC	455	NWSW
235	NGSC	146	NGTV	455	NWSW		
455	NWSW	235	NGSC				
New Zealand		**Russia**		**Italy**		**Thailand**	
Position	**Sequence**	**Position**	**Sequence**	**Position**	**Sequence**	**Position**	**Sequence**
88	NSSL	88	NSSL	1	NISN	23	NQSE
146	NGTV	146	NGTV	21	NSSL	42	NQTY
235	NGCS	455	NWSW	79	NGTI	47	NISN
455	NWSW			168	NGSC	67	NSSL
				319	NFSI	125	NGTI
						214	NGSC
						365	NFSI
Texas (USA)		**Brazil**		**China**		**Kenya**	
Position	**Sequence**	**Position**	**Sequence**	**Position**	**Sequence**	**Position**	**Sequence**
63	NQTY	63	NQTY	63	NQTY	63	NQTY
68	NISN	68	NISN	68	NISN	68	NISN
88	NSSL	88	NSSL	88	NSSL	88	NSSL
146	NGTI	146	NGTI	146	NGTI	146	NGTI
235	NGSC	235	NGSC	235	NGSC	235	NGSC
386	NFSI	386	NFSI	386	NFSI	386	NFSI

Analysis of glycosylation sites

The glycosylation sites (Table-3) identified using NetNGlyc 1.0 (http: //www. cbs. dtu. dk/services/NetNGlyc/) of CBS server, to compare the post-translational modification of the protein sequences. The glycosylation site prediction (Table 3), shows that there is a common glycosylation site at position 146 (NGTV) and 455 (NWSW) in the NA H1N1 sequence from Malaysia, India, USA, New Zealand and Russia. Italy protein sequence of NA H1N1 has a unique glycosylation site i.e. NISN at position 1 which did not exist in any previous appeared strain except USA NA H1N1 strain. This site is present in all the further available sequence after 2009. Sequence from Brazil, China and Kenya holds common glycosylation sites i.e. NQTY, NISN, NSSL, NGTI, NGSC, and NFSI at position 63, 68, 88, 146, 235, 386.

Prediction of antigenic variants

Antigenic variants from all the protein sequences were predicted using CTLPred tool (http: //www. imtech. res. in/raghava/ctlpred/index. html) from the imtech server. It predicts the CTL (Conserved cytotoxic T lymphocytes) epitopes, which helps in the design of the subunit vaccine. The result indicate that position 55-63 (TYENNTWVM), 167-175 (PSPYNSRFE), 228-236 (ESECVCVNG) are conserved in the NA sequences, but 2009 NA H1N sequence shows the unique antigenic variants i.e. SKDNSIRIG, SASACHDGI, and IITDTIKSW at position 33-42, 113-121 and 144-152 respectively.

Phylogenetic analysis

Phylogenetic analysis indicates the separation of one sequence from the other. Its divergence is measured in terms of branch length. The phylogenetic tree indicated that protein Sequences of NA H1N1 from New Zealand, Russia, USA, Malaysia, andIndia areclosely related with theBrazil, China and Kenya with branch length of 0.0606, and sequence of NAH1N1 of Italy and Thailand are distantly related.

Disorder prediction

The disorder region of the protein predicted using PONDR®s VLXT software gives the graphical as well as text view of disorder region (table-5). The threshold value is set to 0.5, for the prediction of disorder region of the sequence. A peak over the threshold value shows the disorder region and those present below the threshold value considered as normal region.

Antigenic variants (table-4) and disorder prediction (table-5) also depicts that there is a similarity between NA H1N1 sequences from Malaysia, India, New Zealand and Russia with the exception of USA. Same way we can say that there is a common similarity among Thailand, Brazil, China and Russia. So from both the tables we can state that Italy NA H1N1 is the unique one among the sequence taken for the study.

Table 4: Antigenic variants for NA H1N1 sequences

Malaysia		**New jersey (USA)**		**India**		**Memphis (USA)**	
Position	**Sequence**	**Position**	**Sequence**	**Position**	**Sequence**	**Position**	**Sequence**
220	ESECVCVNG	220	ESECVCING	228	ESECVCVNG	228	ESECVCVNG
366	SSRKGFEMI	44	SNPKVCNQS	366	SSRKGFEMI	55	TYENNTWVM
55	TYENNTWVN	55	TYENNTWVN	55	TYENNTWVN	167	PSPYNSRFE
167	PSPYNSRFE	167	PSPYNSRFE	167	PSPYNSRFE	235	NGSCFTIMT
179	WASSACNDG	179	WSASACHDG	179	WSASACHDG	239	FTIMTDGPS
New Zealand		**Russia**		**Italy**		**Thailand**	
Position	**Sequence**	**Position**	**Sequence**	**Position**	**Sequence**	**Position**	**Sequence**
206	LTQGALLND	228	ESECVCMNG	34	SKDNSIRIG	35	YENNTWVNQ
230	LMSEPLGEA	242	MTDGPSNGA	113	SASACHDGI	147	SPYNSRFES
377	SFNQNLDYQ	167	PSPYNSKFE	30	WAIYSKDNS	208	SECACVNGS
386	IGYICSGVF	179	WSASACHDG	106	RFESVAWSA	219	TVMTDGPSD
22	ESINFLENA	235	NGSCFTIMT	144	IITDTIKSW	222	TDGPPSDGQA
Texas (USA)		**Brazil**		**China**		**Kenya**	
Position	**Sequence**	**Position**	**Sequence**	**Position**	**Sequence**	**Position**	**Sequence**
42	NQNQIETCN	228	ESECACVNG	42	NQSQIETCN	42	NQSQIETCN
55	TYENNTWVN	42	NQNQIETCN	228	ESECACVNG	220	ESECACVNG
167	PSPYNSRFE	55	TYENNTWVN	55	TYENNTWVN	55	TYENNTWVN
220	ESECACVNG	167	PSPYNSRFE	167	PSPYNSRFE	167	PSPYNSRFE
239	FTIMTDGPS	239	FTIMTDGPS	239	FTIMTDGPS	239	FTIMTDGPS

Table 5: Disorder region of NA sequence

Strain	**Position**	**Disorder**	**No. of disorder**
Malaysia	1-2, 4, 76-82, 148-169, 332-337	MN, N, AGKDTTS, TVKDRSPYRALMSCPIGEAPSPY, KGSCDP	5
New Jersey (USA)	70-89, 148-169, 334-337, 460-464	SNTNIAAGQGVTPIILAGNS, TVKDRSPYRTLMSCPIGEAPSP, NCGP, GADLP	4
India	1-2, 4, 34-37, 79-82, 148-169, 215-224, 332-337, 461-465	MN, N, VSHS, DTTS, TVKDRSPYRALMS CPIGEAPSP, TIKSWRKRIL, KGSCDP, GAELP	8
Memphis (USA)	1-2, 4, 34-38, 80-83, 148-169, 217-224, 329-339, 358-372, 461-465	MN, N, ASHSI, KTSM, TVKDRSPYRALMSCPLGEAPSP, KSWKKRIL, KDGEGSCNPVT, WIGRTKSNRLRKGFE, GAELP	9
New Zealand	1-2, 4, 33-38, 148-165, 217-224, 329-339, 363-371, 461-465	MN, N, WASHSI, TVKDRSPYRALMSCPLGE, KSWKKRI, KDGEGSCNPVT, KSNRLRKGF, GAELP	8
Russia	1-2, 4, 33-38, 148-165, 215, 333-338, 363-370, 461-465	MN, N, WASHSI, TVKDRSPYRALMSCPLGE, T, GSCNPV, KSNRLRKG, GAELP	8
Italy	17-23, 82-100, 265-271, 391-392, 394-396	KLAGNSS, IKDRSPYRTLMSCPIGEVP, TGSCGPV, PD, AEL	5
Thailand	63-69, 128-146, 311-317	KLAGNSS, IKDRSPYRTLMSCPIGEVP, TGSCGPV	3
Texas (USA)	1-2, 4, 84-90, 149-167, 332-338, 461-464	MN, N, KLAGNSS, IKDRSPYRTLMSCPIGEVP, KGSCGPV, AELP	6
Brazil	1-2, 4, 84-89, 149-167, 332-338, 456-464	MN, N, KLAGNS, IKDRSPYRTLMSCPIGEVP, TGSCGPV, SWPDGAELP	6
China	1-2, 4, 84-90, 149-167, 332-338, 461-464	MN, N, KLAGNSS, IKDRSPYRTLMSCPIGEVP, TGSCGPV, AELP	6
Kenya	1-2, 4, 84-90, 149-167, 332-338, 461-464	MN, N, KLAGNSS, IKDRSPYRTLMSCPIGEVP, TGSCGPV, AELP	6

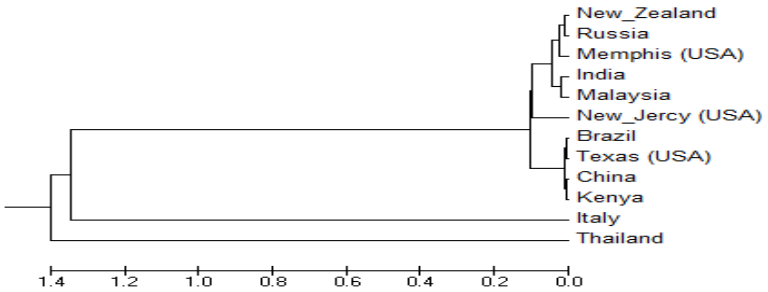

Fig. 1: A phylogenetic tree among various NA sequences obtained from different country

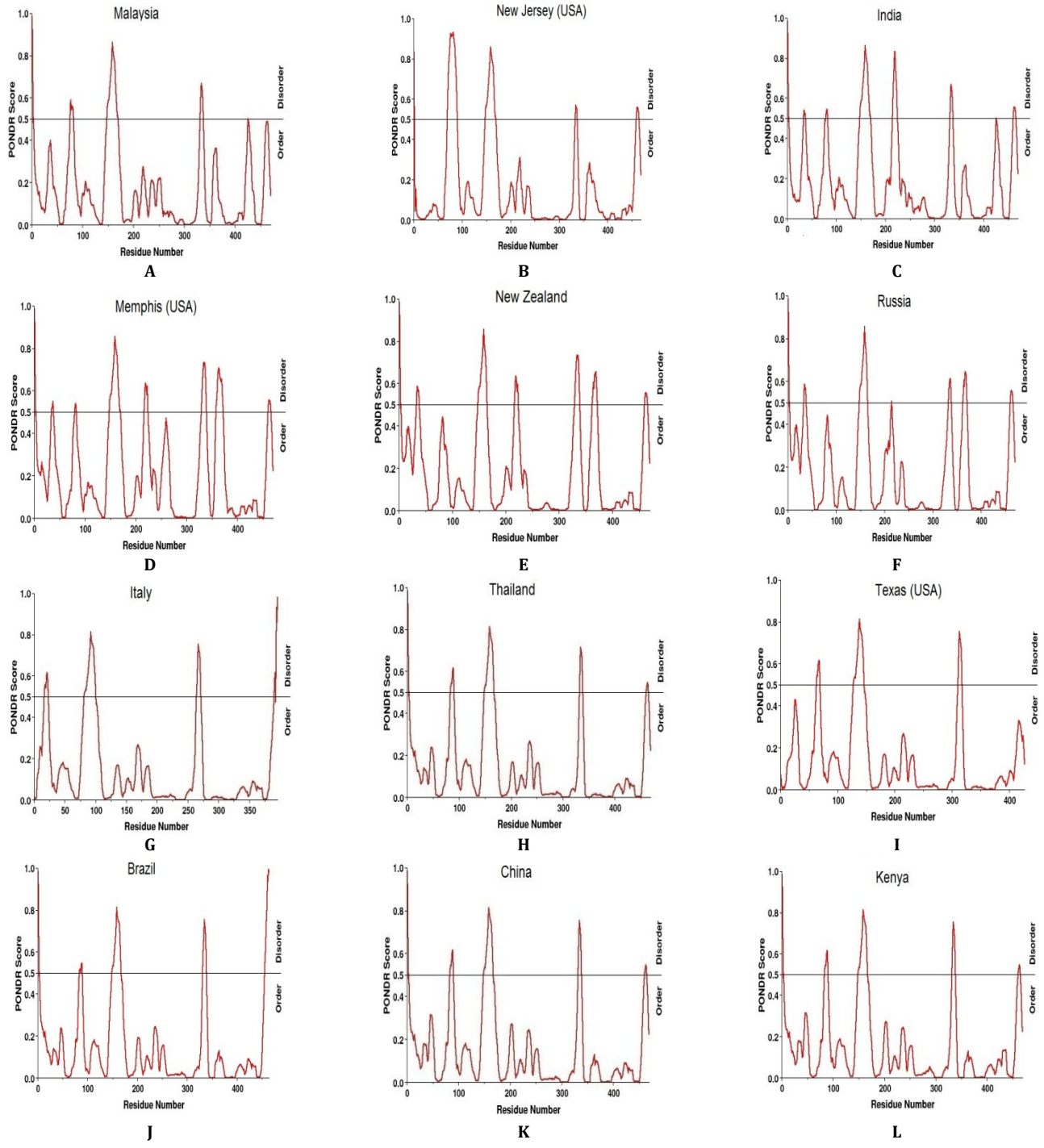

Fig. 2: Graphical representation of Disorder region. The X-axis represents the residue number of the protein sequence, while Y-axis represents the score value. Threshold is the cut-off value for prediction of disorder region

CONCLUSION

The influenza virus enables its spread through the human body by means of its Neuraminidase receptor protein enzyme present on its surface. The NA enzyme facilitates the release and subsequent growth of progeny virions following the intracellular viral replication cycle. NA exhibits its main function during the initial stages of infection when it cleaves sialic acid from the cell surface as well as of the progeny virions, which enable its release from the infected cells and thus it, spreads further into the body by infecting other normal healthy cells [19] Antibodies against the NA enzyme can inhibit it and regulate the infection but the various Antigenic variations of the NA enzyme makes the antibodies ineffective in a vaccine [20].

In this study, we have considered different protein sequences of NA H1N1 strains from different countries to learn about the region of similarity and dissimilarity. So this sequence analysis study revealed that there is a slight difference between these sequences, but the protein sequence of NA H1N1from Italy shows that, it has multiple variations. So our study has found that the protein sequence of NA H1N1 strains from Italy was the unique one. Apart from we also suggest that, protein sequence of NA H1N1 strain from Thailand, Brazil, China and Kenya are similar in characteristics.

CONFLICT OF INTERESTS

Declared None.

REFERENCES

1. Pizzorno A, Bouhy X, Abed Y, Boivin G. Generation and characterization of recombinant pandemic influenza A (H1N1) viruses resistant to neuraminidase inhibitors. J Infect Dis 2011;203(1):25-31.
2. Girard MP, Tam JS, Assossou OM, Kieny MP. The 2009 A (H1N1) influenza virus pandemic: a review. Vaccine 2010;28(31):4895-902.
3. Roll U, Yaari R, Katriel G, Barnea O, Stone L, Mendelson E, *et al*. Onset of a pandemic: characterizing the initial phase of the swine flu (H1N1) epidemic in Israel. BMC Infect Dis 2011;11:1-13.
4. Ahn I, Son HS. Comparative study of the nucleotide bias between the novel H1N1 and H5N1 subtypes of influenza A viruses using bioinformatics techniques. J Microbiol Biotechnol 2010;20(1):63-70.
5. van der Vries E, Collins PJ, Vachieri SG, Xiong X, Liu J, Walker PA, *et al*. H1N1 2009 pandemic influenza virus: resistance of the I223R neuraminidase mutant explained by kinetic and structural analysis. PLoS Pathog 2012;8(9):1-8.
6. Pizzorno A, Bouhy X, Abed Y, Boivin G. Generation and characterization of recombinant pandemic influenza A (H1N1) viruses resistant to neuraminidase inhibitors. J Infect Dis 2011;203(1):25-31.
7. van der Vries E, Veldhuis Kroeze EJ, Stittelaar KJ, Linster M, Van der Linden A, Schrauwen EJ, *et al*. Multidrug resistant 2009 A/H1N1 influenza clinical isolate with a neuraminidase I223R mutation retains its virulence and transmissibility in ferrets. PLoS Pathog 2011;7(9):e1002-276.
8. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. Uni Prot KB/Swiss-Prot. Methods Mol Biol 2007;406:89-112.
9. Apweiler R, Bateman A, Martin MJ, O'Donovan C, Magrane M, Alam-Faruque Y, *et al*. Activities at the universal protein resource (Uni Prot). Nucleic Acids Res 2014;42:191-8.
10. Edgar RC, Batzoglou S. Multiple sequence alignment. Curr Opin Struct Biol 2006;16(3):368-73.
11. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 2000;302(1):205-17.
12. Rice P, Longden I, Bleasby A. EMBOSS: the european molecular biology open software suite. Trends Genet 2000;16(6):276-7.
13. Julenius K, Johansen M, Zhang Y, Brunak S, Gupta R. Prediction of Glycosylation Sites in Proteins. Bioinformatics for Glycobiology and Glycomics; 2010
14. Bhasin M, Raghava GP. Prediction of CTL epitopes using QM, SVM and ANN techniques. Vaccine 2004;22(23-24):195-204.
15. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, andmaximum parsimony methods. Mol Biol Evol 2011;28(10):2731-9.
16. Li X, Romero P, Rani M, Dunker AK, Obradovic Z. Predicting protein disorder for N-, C-, and Internal Regions. Genome Inform Ser Workshop Genome Inform 1999;10:30-40.
17. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins 2001;42(1):38-48.
18. Romero, Obradovic, Dunker K. Sequence data analysis for long disordered regions prediction in the calcineurin family. Genome Inform Ser Workshop Genome Inform 1997;8:110-24.
19. Matrosovich MN, Matrosovich TY, Gray T, Roberts NA, Klenk HD. Neuraminidase is important for the initiation of influenza virus infection in human airway epithelium. J Virol 2004;78(22):12665-7.
20. Air GM, Laver WG. The neuraminidase of influenza virus. Proteins 1989;6(4):341-56.