Consortium of “consistent amino acid substitutions" on Influenza A (H1N1) viral proteome emerged at specific stages of viral infection: A big data analysis

Syeda Lubna and Debashree Bandyopadhyay*

Department of Biological Sciences, Birla Institute of Technology and Science, Pilani, Hyderabad Campus, Hyderabad, Telangana, 500078, India

*Corresponding author; email: banerjee.debi@hyderabad.bits-pilani.ac.in

Abstract:

Influenza-A-virus (IAV) is one of the common threats to humankind since 1918. The viral proteome is frequently substituted leading to new strains and recurrent pandemics. Despite knowing the effects of single amino acid substitutions on individual viral proteins, effects of collective substitutions on viral infection remain elusive. Here, we addressed the question whether the “consistent amino substitutions” occur in consortium on functional domains and protein-protein interaction (PPI) sites, impacting overall viral infection and host immune responses. By definition, "consistent substitutions" occur on “all” the IAV strains isolated in a particular year. Big protein data (563370 sequences and 9824 PPI) and bioinformatics techniques were exploited to address this question. Total one-hundred-and-five "consistent substitutions" were mapped on IAV proteome. Fifty of those emerged on viral functional domains and PPIs, engaged in the specific stages of IAV infection, namely, i) cell surface entry and exit, ii) nuclear import, vRNP assembly and nuclear export and iii) antagonizing immune responses. The study for the first time showed that consortium of “consistent substitutions” emerged on protein functional domains and PPIs, impacting specific stages of viral infection, rather than a single protein, and presumably navigate viral escape from human immune response.

Introduction:

Influenza A virus (IAV) is a human respiratory pathogen that causes recurrent seasonal and pandemics outbreaks across the globe with high rates of morbidity and mortality over the century (1). Four major pandemics were caused by IAV in 1918 (H1N1 Spanish flu), 1957 (H2N2 Asian flu), 1968 (H3N2 Hong Kong flu), and 2009 (H1N1 flu) (2). India recently observed the H1N1 pandemic in 2015, deadlier than in 2009 (3). Despite the availability of drugs and vaccines, the viral escape from the human immune system is likely due to frequent mutations in the segmented genome of the virus, leading to the emergence of new strains (2,4). IAV is one of the rare negative-strand RNA viruses that replicate in the nucleus of the infected cells and utilizes the maximum number of host cellular machinery to replicate and transcribe the viral proteins. Ten viral proteins (that constitutes the viral proteome) were produced by eight segments of viral genome, namely, surface proteins (HA and NA), matrix proteins (M1 and M2), non-structural proteins 1 and 2 (NS1 and NS2, also known as nuclear export protein (NEP)), polymerase complex proteins (polymerase basic-1 protein (PB1), polymerase basic-2 protein (PB2), and polymerase acidic protein (PA)), and nucleoprotein (NP). The maximum number of genetic mutations (and amino acid substitutions) were observed in the surface glycoproteins of IAV, namely hemagglutinin (HA) and neuraminidase (NA), which are responsible for viral entry and proliferation (5). However, the virus adapts multiple substitutions in different segments of the viral proteome that affect virulence. For example, 11 substitutions across the viral proteome in a mice model showed the conversion of an avirulent viral strain to a virulent strain that killed the mice at a minimal viral dose (6). The viral infection cycle (viral entry into the cell, viral replication, transport and proliferation) operates through multiple viral protein complexes and host machineries (Figure 1). Steps of the infection cycle are, attachment of the viral HA glycoprotein to the cell surface, M2 mediated entry, release of viral RNP to cytoplasm, nuclear import of vRNP complex, transcription of new viral RNA, formation of new viral particles, nuclear phenomena mediated by polymerase complex (PB1, PB2 and PA), NP, NEP and M1 proteins, release of new viral particles to the cytoplasm (7,8), expression of NS1 protein to suppress host immune responses (8), and, finally, budding of new virions using M1, M2, and NA proteins (8,9). Substitutions on functional domains of any of these proteins involved in the infection cycle are likely, to disrupt the viral activity and host responses.

So far, no computational attempt is known, to the best of our knowledge, that investigates multiple amino acid substitutions across the influenza viral proteome and their possible consequences on the viral infection cycle.

Here we studied amino acid substitutions (those were consistent over the years) on the influenza A(H1N1) pdm09 viral proteome using multiple sequence alignment and their analyses. Almost half of the consistent substitutions were located on the functional domains of the viral proteins, as obtained from different databases. The role of these functional domains, curated from PubMed, and the effects of these consistent substitutions are discussed. A few "consistent substitutions" were identified that were not involved in functional domains, structure, or function. These substitutions are either not reported experimentally/clinically or could be silent substitutions that escape recognition by a host.

Methodology:

1. **Generation of influenza A(H1N1) pdm09 Proteome dataset:**

The complete proteome of the Influenza A(H1N1) pdm09 strain consists of ten proteins, namely, hemagglutinin (HA), neuraminidase (NA), matrix protein 1 (M1), matrix protein 2 (M2), nucleoprotein (NP), non-structural protein 1 (NS1), nuclear export protein (NEP, synonymous to NS2), and three polymerase subunits (polymerase acidic protein (PA), polymerase basic protein 1 (PB1), and polymerase base protein 2 (PB2)). Sequences of all these ten proteins were retrieved from the NCBI Influenza Resource Database (Influenza virus database – NCBI (nih.gov)) and the Global Initiative on Sharing All Influenza Data (GISAID Initiative (epicov.org)) databases for Indian and Global strains.

1.1. Indian data set for influenza A(H1N1) pdm09 virus:

NCBI Influenza Virus Resource was searched with the following parameters, (i) Type – 'A' (ii) Host – 'human' (iii) Country/Region – 'India' (iv) Subtype – 'H1' or 'N1' (v) Protein – each protein, selected individually, (vi) Collection date – from 2009 to 2022, and (vii) 'Full-length only.' Selection criteria used for the GISAID database are as follows: (i) Dataset – 'EpiFlu™,' (ii) Type – 'A' (iii) Strain – 'H1' and 'N1', (iv) Lineage – 'pdm09' and 'unknown' (vi) Host – 'Human' (vii) Location – 'Asia' and 'India' (viii) Collection date – from '2009' to '2022', (ix) Required segment – individual protein name and (x) 'Only complete.' The data retrieval in this work was done on December 2022. A total of 6,633 sequences were curated for the "Indian dataset" (Table S1).

1.2 ***Global data set for influenza A(H1N1) pdm09 virus:***

The global data set was curated using similar search criteria as the Indian dataset, except the following, a) Country/Region – 'any' in the NCBI database, and b) Location – 'all' in the GISAID database. A total of 5,63,370 sequences were curated for the "Global dataset" (Table S1). Note Indian dataset is a subset of the Global dataset.

2. **Multiple sequence alignment of Indian and Global datasets:**

Multiple sequence alignment (MSA) was performed on ten different protein sequences from a) the Indian and b) Global datasets. A total of twenty different MSA was obtained. ClustalW (10) tool in MEGAX software (11) was employed to obtain the MSA. Following parameters were used in

ClustalW: (i) Pairwise alignment with Gap Opening Penalty: '10' and Gap Extension Penalty: '0.10', (ii) Multiple alignments with Gap Opening Penalty:' 10.00', and Gap Extension Penalty: '0.20', (iii) Negative matrix weight: 'OFF' (iv) Delay divergence cut-off: '30%', and (v) 'Use Predefined Gap' was not selected.

2.1 Pre-processing of MSA data to derive consistent substitutions:

MSA data obtained from MEGAX software were further analysed using Microsoft Excel. There were three components of MSA data (converted to three columns in Excel), namely, (i) sample id, (ii) year of sample collection, and (iii) amino acid sequence. The data were sorted according to the 'year of sample collection' column using Excel's "sort" command. The amino acid sequence variation per sequence position was noted within each block of a year. If an amino acid substitution remained same across the strains throughout a year, that particular substitution was considered as a "consistent substitution." Given the limitation, the sampling might be incomplete due to the non-availability of sequences from all possible strains; for example, the patient sample sequences not analysed or not included in the repository, etc.

2.2 Influenza A(H1N1) virus interaction study:

Protein-protein interactions from the H1N1 virus and human host, along with their interaction sites, were curated from three different databases, namely, Protein Data Bank (RCSB PDB: Homepage), Human – Virus Interaction database (HVIDB-Home (zzdlab.com)), and PubMed (PubMed (nih.gov)).

a) Protein-protein interactions curated from the Human-Virus Interaction database (HVIDB):

HVIDB is a comprehensive database containing 48,644 experimentally verified protein-protein interactions (PPI) from 35 virus families (12). The entire HVIDB dataset containing 48,644 human-virus PPIs was downloaded and filtered for influenza A (H1N1) virus only, which resulted in 9,824 PPIs involving Influenza A (H1N1) virus and human host (Figure 2).

b) Protein-protein interactions curated from Protein Data Bank (PDB):

Structure-based viral-human host protein interaction was studied using the protein complexes in the Protein Data Bank (PDB) database. The search criteria in the PDB database was "H1N1", which resulted in a total of 524 structures. These structures were filtered to remove non-protein complex structures, like structures of antibodies. The final number of entries from the PDB search result was 449 (Figure 2).

c) Details of protein-protein interaction regions from in-vitro and in-vivo experiments:

Details of H1N1 viral protein-host interacting partners and the interaction sites (amino acid residues) were extracted from the PubMed database. The search criteria used were "influenza A virus

[Title/Abstract] AND "H1N1" AND "interaction" AND "residue." The search resulted in 79 articles. Four were excluded those involved comparison between IAV and others, namely, SARS-CoV-2, Alzheimer, and the anti-viral activity of amphibian secretions. As no review articles were obtained from the above search criteria, additional search criteria – "Influenza Avirus–host protein interactions" along with two filters: i) year of publication: "2020-2022", and ii) article type: "review" were used separately, that resulted into eight review articles.

Results and Discussions:

Chronological emergence of "consistent substitutions" in the viral proteome of IAV pdm09 strains isolated from India and the world:

Within the time-period 2009-2022, 105 "consistent substitutions" were observed in the IAV pdm09 proteome (consisting of 4,454 residues) isolated from India. The chronological emergence of "consistent substitutions" in individual proteins was shown [Figure 3a and Table S2]. As defined in the method section, the "consistent substitutions" remained invariant at a particular position across all the strains in the years following its emergence [Figure 3a]. These "consistent substitutions" were distributed across the proteome independent of the protein length. Maximum number of "consistent substitutions" were observed in HA (29), followed by NA (26). The substitutions in HA are predominantly hydrophilic (22 out of 29), whereas the amino acid substitutions in NA are both hydrophobic and hydrophilic in nature. The substitutions in PB2 are predominantly hydrophobic in nature (7 out of 13). Thus, the nature of the substitutions was presumably specific to viral proteins [Figure 3a].

In 2021, there was a sudden hike in the number of "consistent substitutions" throughout the viral proteome [Figure 3b]. Since 2020, a total of 30 "consistent substitutions" were observed. As expected, the surface proteins HA and NA showed maximum substitutions. Apart from those, PA protein has also encountered drastic changes (five consistent substitutions) in 2021, in contrast those (two) before 2015 (post-2009 pandemic-era in India).

To understand whether these "consistent substitutions" were geography specific, we extended the study to the global dataset [Table S3]. The global dataset was manifold larger than the Indian dataset. The substitutions observed in the Indian dataset were, more or less, observed in the Global dataset. However, due to the larger sample size, the substitutions were not always "consistent"; that is, the number of times the substitutions were observed was lesser than the number of strains in the dataset. We have computed "percent consistent substitutions" for global datasets, which indicates the deviation from 100% consistent values. The "percent consistent substitutions" were marginally lower than the "consistent substitutions" in many of the cases (Table S3). The notable observation was, despite the larger sample size, some of these substitutions became "consistent" in 2022, for example, 8 in PB2, 1 in PB1, 2 in PA, 2 in HA, 1 in NA, 6 in NS1, and 3 in NEP. The comparative study

between the Indian and the global dataset indicated similar trends in the "consistent substitutions"; the difference observed was mostly due to an increase in the sample size.

Some of these "consistent substitutions" (HA:K180Q, NS1:E125D, NS1:N205S) were studied earlier as a single mutation on individual proteins (13–15). The engineered single substitutions reported in the literature were listed (Table S4). Loci of 105 “consistent substitutions” on the viral proteome were noted in this study (Figure S1).

50 out of 105 "consistent substitutions", occurred on the functional domains of the viral proteins or on PPIs (as evident from crystallographic, experimental, and clinical data) involving human host and influenza virus (Figure 4).

Mapping of "consistent substitutions" on viral protein functional domains:

"Consistent substitutions" were mapped on four different viral functions, namely, i) nuclear transport signals, ii) receptor binding of HA, iii) antigenicity of HA, and iv) post-translational modifications. The protein domains associated with these viral functions and the "consistent substitutions" evolved therein are discussed in the following subsections.

Nuclear transport signal motifs: 16 nuclear transport signal motifs are present in seven viral proteins (PA, PB1, PB2, NP, M1, NS1, and NEP). There are three different types of nuclear transport signal motifs, namely, nuclear localisation signal (NLS), nuclear accumulation signal (NAS), and nuclear export signal (NES) (8). These motifs are mostly conserved, except for the NLS motif of PA, PB1, and PB2 that showed four "consistent substitutions" (PB1:V200I, PB2:S453T, PA:S225C, and PA:L226I). Two of these four substitutions (PA:S225C and L226I) emerged in 2021; no mutational studies have been reported so far [Table 1a].

Receptor binding sites on HA protein: There are three receptor binding sites (RBS) on the globular head domain of HA where sialic acid attaches (2) at residue positions a) 144 to 154 (130-loop), b) 204 to 212 (190-helix), and c) 234 to 244 (220-loop). As sialic acid binding to HA protein is an important step to viral entry, a significant substitution, reported earlier, D239N (numbering according to H1 strain), impacted glycan binding specificity (3,16,17); however, that was not observed in our present "consistent substitutions." There are four “consistent substitutions” in RBS, i) 130-loop: N146D, ii) 130-loop: K147N, iii) 190-helix: Q206E, and iv) 220-loop: E241A [Table 1b].

Antigenic sites on HA protein: There are four antigenic sites, Sa, Sb, Ca, and Cb, on the HA protein of H1 strains (18). As HA protein makes the initial contact with the host cells via binding to sialic acid, frequent substitutions occur on HA antigenic sites (18) that most likely escape the human immune system. A total of nine "consistent substitutions" were observed in these four antigenic regions from 2014 to 2022 [Table 1c]; a) one in 2014 (before the post-2009 pandemic in 2015), b) two in 2017, c) one in 2018, d) one in 2020, and e) four in 2021. The maximum changes are observed

in the antigenic site, Sa, (i) K180Q in 2014, ii) S179N in 2017, iii) S181T in 2018, and iv) N173K, and L178I in 2021) whereas the antigenic site Ca is conserved [Figure S2].

Post-translational modification sites: Five post-translational modifications were reported in different viral proteins, namely, phosphorylation (13,19–22), ubiquitination (21), palmitoylation (23), SUMOylation (24), and glycosylation (25,26). Out of these five, except glycosylation, all were conserved throughout the proteome of IAV pdm09 strains [Table S5]. Glycosylation sites on HA and NA proteins exhibited "consistent substitutions" (Table 1d).

"Consistent substitutions" at the glycosylation sites along the proteome:

HA protein: Potential asparagine glycosylation sites on HA protein from Indian sequences (time period 2009 to 2022) were identified using NetNGlyc 1.0 (27). The HA protein sequences from the Indian dataset were subject to redundancy removal (100% sequence identity) using CD-HIT (28). Three representative sequences per year (the minimum number available for a year) were selected from the non-redundant HA sequences. Those sequences were scanned to identify the potential N-glycosylation sites using the NetNGlyc 1.0. The prediction results were recorded only for the "Asn-Xaa-Ser/Thr" sequons (27,29) and those were in agreement with experimental observations (29). Five conserved glycosylation sites (^{40}NVTV^{43}, ^{104}NGTC^{107}, ^{304}NSSL^{307}, ^{498}NGTY^{501}, and ^{557}NGSL^{560}) were identified. One new glycosylation site emerged in 2017 (^{179}NQSY^{182}) due to the "consistent substitution" S179N (Table 2a). "Consistent substitution" S181T might have affected the glycosylation at site ^{179}NQTY^{182}, as it was shown that the NXT glycosylation motif was preferred over NXS (26). Note that these two "consistent substitutions" (S179N and S181T) were also present on the antigenic region of HA.

NA protein: Similar to HA protein, potential asparagine glycosylation sites on NA proteins were identified using the above-mentioned method, resulting in nine potential glycosylation sites. Five conserved glycosylation motifs (^{58}NNTW^{61}, ^{68}NISN^{71}, ^{88}NSSL^{91}, ^{146}NGTI^{149}, and ^{235}NGSC^{238}) were identified on NA protein (time period 2009-2022). Knock-down studies suggested an important role of glycosylation at Asn^{235} in the budding, replication, and virulence of the H1N1 influenza virus (25). Two of the "consistent substitutions" (Q51K and Y66F) were mapped on two glycosylation motifs (^{50}NQSV^{53} and ^{63}NQTY^{66}) [Table 2b]. A third glycosylation motif was abolished in 2015 (^{386}NFSI^{389}) due to the "consistent substitution" N386K. The fourth one emerged (^{42}NQSQ^{45}) in the same year due to the "consistent substitutions" of N44S.

These observations indicated that glycosylation sites in IAV pdm09 strains are under intense selective pressure, also reported earlier (18,26).

Mapping of the "consistent substitutions" on protein-protein interactions (PPIs):

Protein-protein interactions are crucial in mediating the replication of the IAV in the nucleus of the infected cells, transcription of the viral proteins, and final budding of the new virus particles out of

the host cell (Figure 1). The viral proteins communicate among themselves to complete the infection cycle. Hence, two types of protein-protein interactions (PPIs) are important, i) human host and viral protein interactions (host-viral PPI) and ii) viral protein-protein interactions (vPPIs). Effects of the "consistent substitutions" on both types of PPIs, if any, were extracted from three databases, PubMed, PDB, and HVIDB. There were 9,824 interacting partners extracted from HVIDB. No information was available on interaction sites from HVIDB. Forty-five experimental interaction sites were obtained from PubMed and PDB databases. Of these forty-five, twenty-two sites were conserved (Table S6), and twenty-three sites hosted “consistent substitutions” (Table 3).

"Consistent substitutions" in host–viral PPI:

33 **host** interacting factors with experimentally known interaction sites were manually curated from PubMed and PDB. Of the 33 known interaction sites, 18 were conserved [Table S6a]. The remaining 15 interaction sites were non-conserved and exhibited "consistent substitutions" [Table 3a]. The maximum number of "consistent substitutions," nine, were observed on the NS1 protein sequence at the host factor binding sites. Most likely, this is because NS1 is a multifunctional protein (20,30). NS1 protein has two major functional domains, RNA binding domain (RBD) and effector domain (ED). The "consistent substitutions" on RBD (D2E, E55K) are less in number, presumably because it binds to fewer host factors, mainly RNA. The "consistent substitution" (T80A) is present in the linker region and just before the interaction sites of eIF4GI (between residues 81-113) and PABPII (at residue 81). There are six "consistent substitutions" (V84I, L90I, E125D, K131E, A155T, and N205S) in ED. These “consistent substitutions” of ED involve multiple host factor interactions that may intervene with multiple functions, such as nuclear export, viral replication, cell apoptosis, etc [Table 3a]. Eight of these nine consistent substitutions on NS1 protein were reported in our previous study, except V84I, which emerged in 2021 (30). Considering the frequent substitutions of NS1 protein, it was proposed as a possible drug target for IAV (31–33).

Apart from NS1, a few more viral proteins exhibited "consistent substitutions" at the host factor binding sites. For example, eight "consistent substitutions" (M1:V80I, M1:M192V, NP:A22T, NP:V425I, NP:S498N, NEP:D2E, NEP:N29S, NEP:T48A) may have an impact on nuclear export of different viral proteins as these are present on the host factor binding sites involved in nuclear export. Another "consistent substitution" (M1:V80I) may affect the inhibitory activities of viral proteins interacting with the host factors involved in the host immune response (34). "Consistent substitution" (NP:A22T) may affect the transcription of the viral genome, as it is present at interaction sites involved in binding to the inhibitory factor, MxA (35).

A total of seventeen "consistent substitutions" were observed on fifteen sites involving human host and viral protein-protein interactions.

There were 20 additional host interacting partners reported with unknown interaction sites [Table S7]. Due to the lack of information, the possible effects of “consistent substitutions” on the host factors are unknown.

"Consistent substitutions" in viral-viral PPIs (vPPIs):

Fourteen vPPIs were extracted from different databases, twelve of which have known interaction sites – four were conserved (Table S6b), and eight were non-conserved. A total of seventeen "consistent substitutions" (PB2:R54K, PB2:M66I, PB2:T81I, PB2:D195N, PB2:G225S, PB2:V667I, PB2:V731I, PB1: G154D, PB1:V200I, PB1:K386R, PB1:I397M, PB1:I435T, NP:A22T, NP:V425I, NP:S498N, M1:V80I and NEP:M83I) were mapped on eight vPPI sites [Table 3b]. Although NS1 is involved in multiple host factor binding, it is absent in vPPIs.

The viral polymerase complex has multiple vPPIs from three interacting viral proteins, namely, PB2, PA, and PB1. The PB1-PA interaction sites were conserved, but the interaction sites involving PB2 showed seven “consistent substitutions” (R54K, M66I, T81I, D195N, G225S, V667I, and V731I). These substitutions may impact the endonuclease activity of PB2 and subsequent viral replication and transcription processes (36,37). Three "consistent substitutions" (K386R, I397M, and I435T) on the PB1-cRNA (viral) interaction site may affect the viral replication process (38).

The heterotrimeric (PA-PB1-PB2) viral polymerase complex interacts with the NP-vRNA complex to form the viral ribonucleoprotein complex (vRNP) (Figure 1). Twelve "consistent substitutions" (PB2:R54K, PB2:M66I, PB2:T81I, PB2:D195N, PB2:G225S, PB2:V667I, PB2:V731I, PB1:G1154D, PB1:V200I, NP:A22T, NP:V425I, and NP:S498N) were observed in vRNP complex interaction sites (protein 1 – protein 2 in Table 3b: PB2-PB1, NP-vRNA, NP-PB2, NP-PB1, NP-NP). Thus, these substitutions may impair the vRNP complex formation and functioning.

The V80I "consistent substitution" is present on the M1-M1 dimer interface that constitutes a layer underneath the viral membrane, holding vRNP on one side of the layer and viral membrane proteins (HA, NA, and M2) on the other side (8). This substitution (V80I) on M1, along with M83I substitution on NEP, may impact the nuclear export of the M1-vRNP complex [Table 3b].

To note, four "consistent substitutions," one on M1 (V80I) and three on NP (A22T, V425I, and S498N), are mapped on both host-viral and vPPIs [Table 3].

Consortium of “consistent amino acid substitutions” emerged at the specific stages of IAV infection cycle and facilitated escape from human immune response:

The “consistent substitutions” were mapped on the viral protein domains and protein-protein interaction sites involving human host and viral factors, based on the experimental, mutational and clinical observations. The results showed that there were total forty-two conserved sites on the viral proteome (twenty on post-translational modifications, eighteen on human-viral PPIs and four on viral-

viral PPIs). In contrast, twenty-four functional domains on the viral proteomes hosted thirty-one "consistent substitutions". The "consistent substitutions" were mainly observed at the receptor binding sites and antigenic sites of the HA protein (during viral entry to the host cell), glycosylation sites of HA and NA proteins (attachment points of the virus to the host cell membrane either during viral entry or budding of new viral progeny) or at the nuclear localization signals of the viral polymerase complexes (PA, PB1 and PB2 proteins), involved in nuclear import of vRNPs (Figure 4).

In addition, there were thirty-four protein-protein interaction sites (involving human-virus and viral protein complexes) those hosted twenty-three "consistent substitutions". Nine "consistent substitutions" were located on NS1 protein only (Figure 4). NS1 protein is expressed at the later stage of the infection and is involved in multiple functions. The "consistent substitutions" on NS1 protein are involved in following host factor binding (PPIs) mainly suppressing immune response, blocking nuclear processes and enhancing viral replication and transcription. Those host factors are - i) the cellular dsRNA, ii) the sensors, those activate innate immune responses, namely, interferon-inducible RNA-dependent protein kinase (PKR), retinoic acid-inducible gene-I proteins (RIG-I), TRIM25-mediated RIG-I activation and 2'-5'-oligoadenylate synthetase (OAS/RNaseL), iii) the poly(A)-binding protein II (PABPII) iv) the nuclear RNA export factor 1 (NXF1), v) the cleavage and polyadenylation specificity factor 30 (CPSF30), which prevents CPSF30 from binding cellular pre-mRNAs, vi) phosphoinositide 3-kinase (PI3K) that activates the AKT pathway (involved in delaying virus-induced apoptosis and facilitating viral replication NS1), vii) CT-10 regulator of kinase (Crk and/or CrkL) that activates PI3K, viii) eukaryotic translation initiation factor 4GI (eIF4GI) and PABPI that enhances viral mRNA translation. In contrast, "consistent substitutions" were observed on NP protein at host-viral PPI (MxA-NP) site those suppress viral mRNA transcription (facilitating host-defence mechanism).

Other than NS1, M1 is another multifunctional protein. The "consistent substitutions" emerged on M1 protein – host factor binding sites (PPIs) those are responsible for nuclear export of vRNPs (HSP70-M1 interaction) and suppression of classical complement pathway (C1qA-M1 interaction). Nuclear export of vRNPs and viral proteins are also facilitated by other PPIs, namely, NTF-2-NXT-1-NP, hNup98-NEP, and PKIα-NEP, those hosted multiple "consistent substitutions".

The consortium of "consistent substitutions" emerged on the binding sites of various viral protein complexes (PPIs), involved in vRNP assembly (NP-PB1, NP-NP, NP- PB2, PB1-PB2, and vRNA-NP), viral assembly (M1 dimerization), viral replication (PB1-cRNA), and nuclear export of vRNPs((M1-vRNP)-NEP) (Figure 5).

Based on the current report following inferences can be drawn – the "consistent substitutions" emerged on the viral protein domains and PPIs those are i) either at the viral entry or exit points on the cell surface, ii) involved in nuclear events, like, nuclear import, replication and transcript, viral assembly and nuclear export (to note IAV is a negative stranded RNA virus where the virus hijacks most of the host machineries for its multiplication), iii) involved in alteration of host-immune responses. Thus, it seems that the "consistent substitutions" emerging on the specific stages of the viral infection, act as navigator to the virus directing it towards an alternate direction (such as, emergence of new viral strain) when encountered by host immune responses.

Conclusion:

Influenza A Virus is a threat to the human kind since the last one hundred years. The frequent substitutions on the proteome of this negative-sense RNA virus lead to ineffectiveness of the existing drugs and introduction of new vaccines, recurrently. Understanding of the substitution patterns across the viral proteome would presumably highlight whether the single "substitutions" affect individual protein functions or consortium of those substitutions on a pool of viral proteins affect the overall viral functions and host immune responses. The literature reports were limited to the amino acid substitution studied on a single viral protein. Here we studied the proteome-wide (all ten viral proteins) amino acid substitutions those became "consistent" over the years, using multiple sequence analysis. Three of the "consistent substitutions" (NS1: N205S, NS1:E125D and HA: K180Q) reported here were studied earlier as engineered single substitutions on individual proteins. According to the earlier studies, two of these substitutions – i) NS1:N205S was involved in NS1-mediated enhancement of viral polymerase activity and ii) NS1:E125D was involved in CPSF30 binding that blocks the processing of host cellular mRNAs thus antagonise host innate immune responses. The current study identified that these two substitutions become "consistent" since 2014 (NS1:N205S) and 2016 (NS1:E125D). However, these substitutions were identified as a part of a consortium of "consistent substitutions" on NS1 protein (total 9), involved in antagonising human immune response. NS1:E125D, as a part of the consortium of "consistent substitutions", might affect PABP1 and PABPII binding, TRIM25 mediated IFN response, RIG-I CARD ubiquitination, and host interferon system. Similarly, NS1:N205S, as a part of the consortium of "consistent substitutions", might interfere with multiple host factors, namely, NXF1 complex, PI3K, CRK/CRKL, PABP1/PABPII, TRIM25 and PKR.

Mapping of the "consistent substitutions" across the viral proteome was done for the first time. Consortium of "consistent substitutions" were identified on multiple protein functional domains and PPIs those evolved at the specific stages of the viral infection, namely, i) entry and exit points on the cell surface, ii) entry and exit point on the nuclear surface and iii) involvement in antagonizing human immune response. Emergence of consortium of "consistent substitutions" at the specific stages of viral infection on human host presumably indicates the roadmap for the virus to escape the human

immune response leading to new strains (might be with sustained on enhanced virulence). The postulate is subject to further experimental verifications. Some of the "consistent substitutions" were involved in PPIs engaging both host-viral and viral-viral interactions. However, NS1 protein exclusively participates in human-viral interaction. Listing of the conserved and non-conserved sites on the viral proteomes opened an avenue to explore the drugability of the interaction sites, subject to availability of the three-dimensional structures. The druggable pockets can be screened through the drug libraries that would facilitate efficient repurposing of the existing drug (or drug-like) molecules.

Figure 1: Infection cycle of H1N1 influenza A virus involving different host factors and viral proteins. Viral proteins involved in the infection cycle are shown along the arrows. Infection cycle steps: - 1) HA attachment, 2) acidification of viral interior, 3) fusion of viral ad endosomal membranes, 4) release of vRNPs, 5) nuclear import of vRNPs, 6) transcription and replication of viral RNA, 7) vRNP assembly, 8a) nuclear export of viral mRNA to endoplasmic reticulum, 8b) nuclear export of vRNPs to cytoplasm, 9a) transport of newly synthesized viral proteins (except NS1) to the cell membrane, 9b) transport of vRNPs to the cell membrane, 10) packaging and budding of new virions

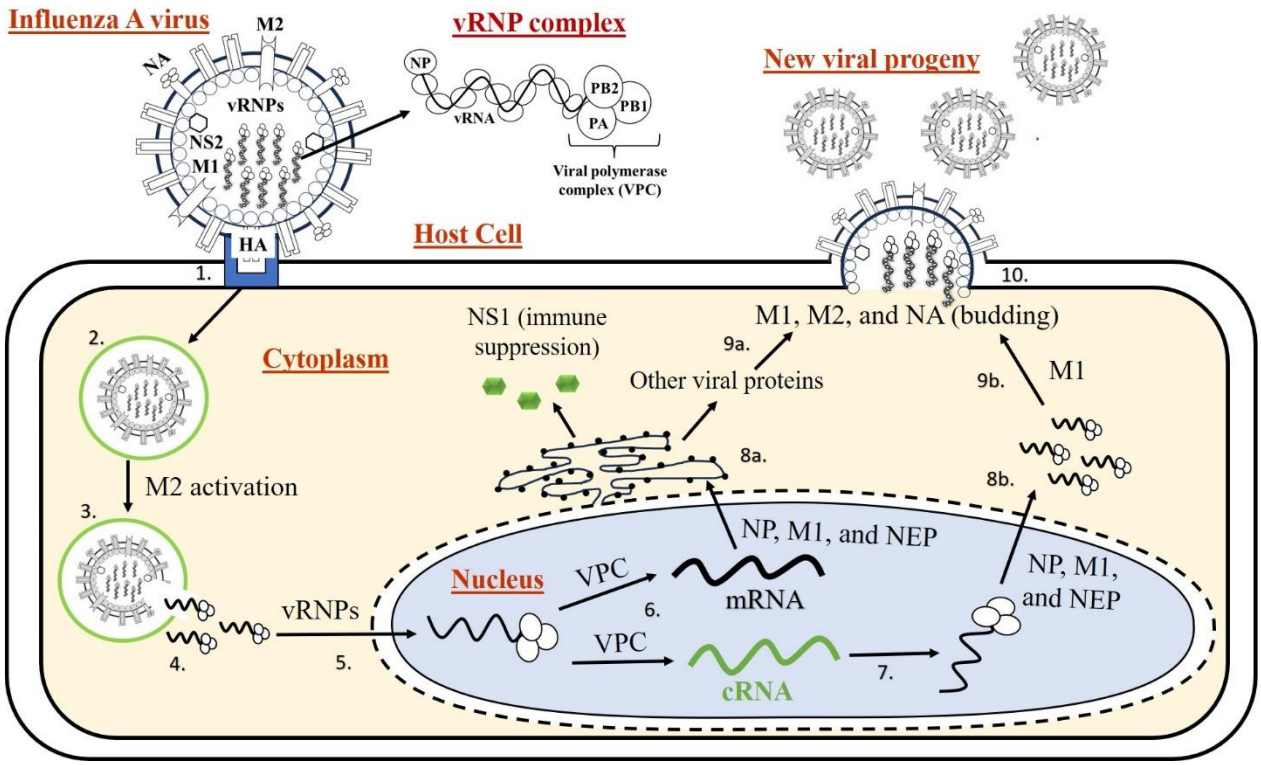
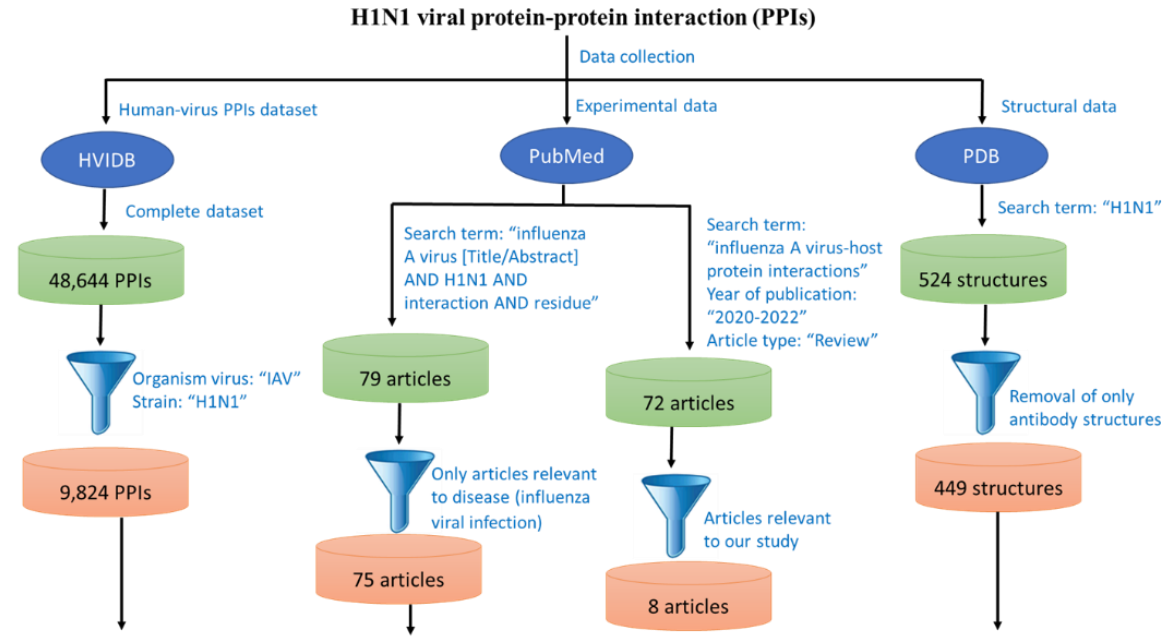


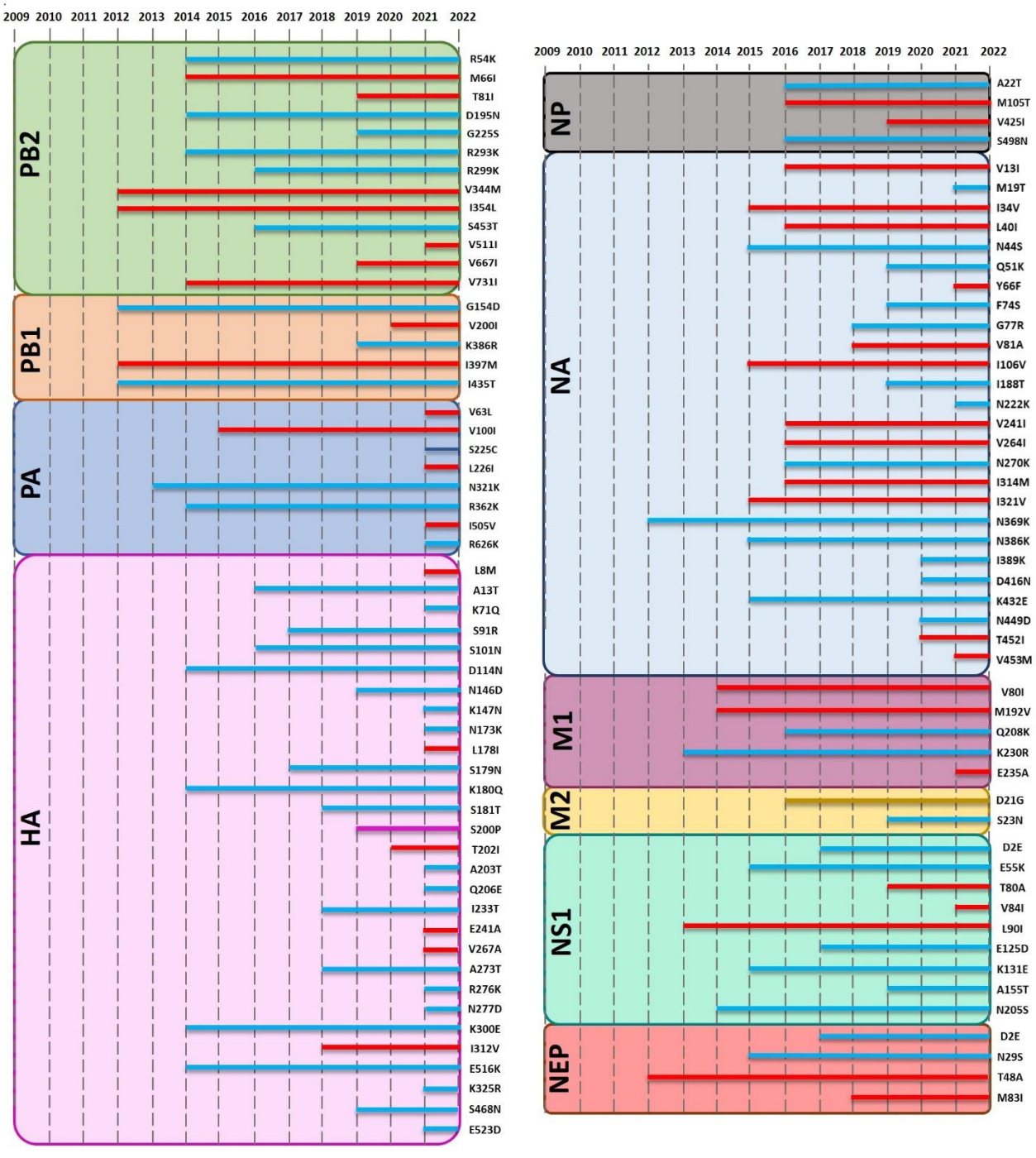
Figure 2: Human-H1N1 viral protein-protein interactions derived from different sources



Protein Name	No. of HVIDB PPIs	No. of PubMed articles	No. of PDB structures (PPIs)
PB2	1,432	9	180 (18)
PB1	1,197	4	
PA	621	3	
HA	687	17	117 (1)
NP	1,096	5	22 (9)
NA	1,090	9	32 (0)
M1	796	5	21 (3)
M2	1,441	4	7 (0)
NS1	1,081	17	14 (7)
NEP	383	2	1 (0)
Total entries	**9,824**	**75**	**449 (38)**

Figure 3: Consistent substitutions in 10 viral proteins (whole proteome) in the Indian dataset a) per viral protein, appearance and retention over time and b) variation in total number over time. The colored box represents individual proteins.

a)


2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022
PB2
R54K
M66I
T81I
D195N
G225S
R293K
R299K
V344M
I354L
S453T
V511I
V667I
V731I
PB1
G154D
V200I
K386R
I397M
I435T
PA
V63L
V100I
S225C
L226I
N321K
R362K
I505V
R626K
HA
L8M
A13T
K71Q
S91R
S101N
D114N
N146D
K147N
N173K
L178I
S179N
K180Q
S181T
S200P
T202I
A203T
Q206E
I233T
E241A
V267A
A273T
R276K
N277D
K300E
I312V
E516K
K325R
S468N
E523D
2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022
NP
A22T
M105T
V425I
S498N
NA
V13I
M19T
I34V
L40I
N44S
Q51K
Y66F
F74S
G77R
V81A
I106V
I188T
N222K
V241I
V264I
N270K
I314M
I321V
N369K
N386K
I389K
D416N
K432E
N449D
T452I
V453M
M1
V80I
M192V
Q208K
K230R
E235A
M2
D21G
S23N
NS1
D2E
E55K
T80A
V84I
L90I
E125D
K131E
A155T
N205S
NEP
D2E
N29S
T48A
M83I

b)

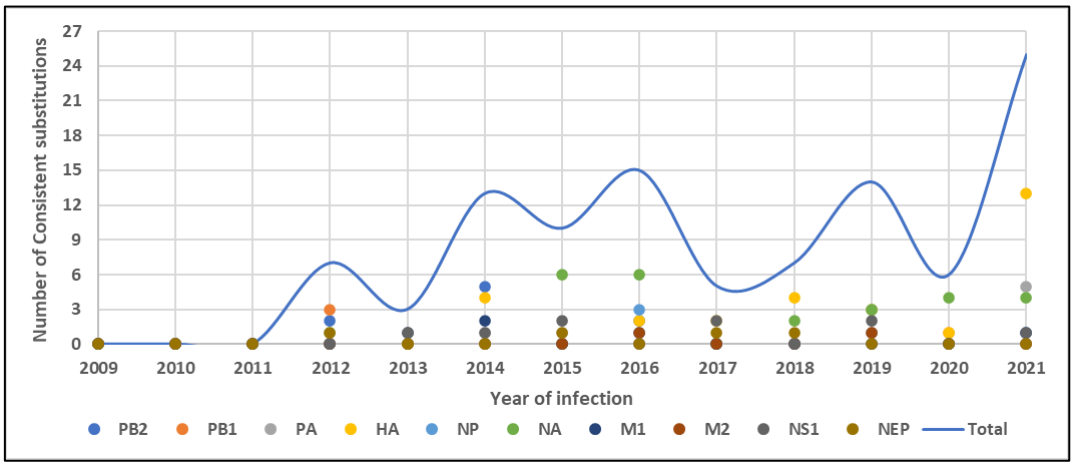

Number of Consistent substitutions
27
24
21
18
15
12
9
6
3
0
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Year of infection
PB2
PB1
PA
HA
NP
NA
M1
M2
NS1
NEP
Total

Figure 4: 50 "Consistent substitutions" on IAV proteome involving the a) viral functions, b) host-viral protein-protein interactions and c) viral – viral protein-protein interactions. 4 "consistent substitutions" were common across the functional domains and PPIs. The proteins are arranged according to their sequence lengths

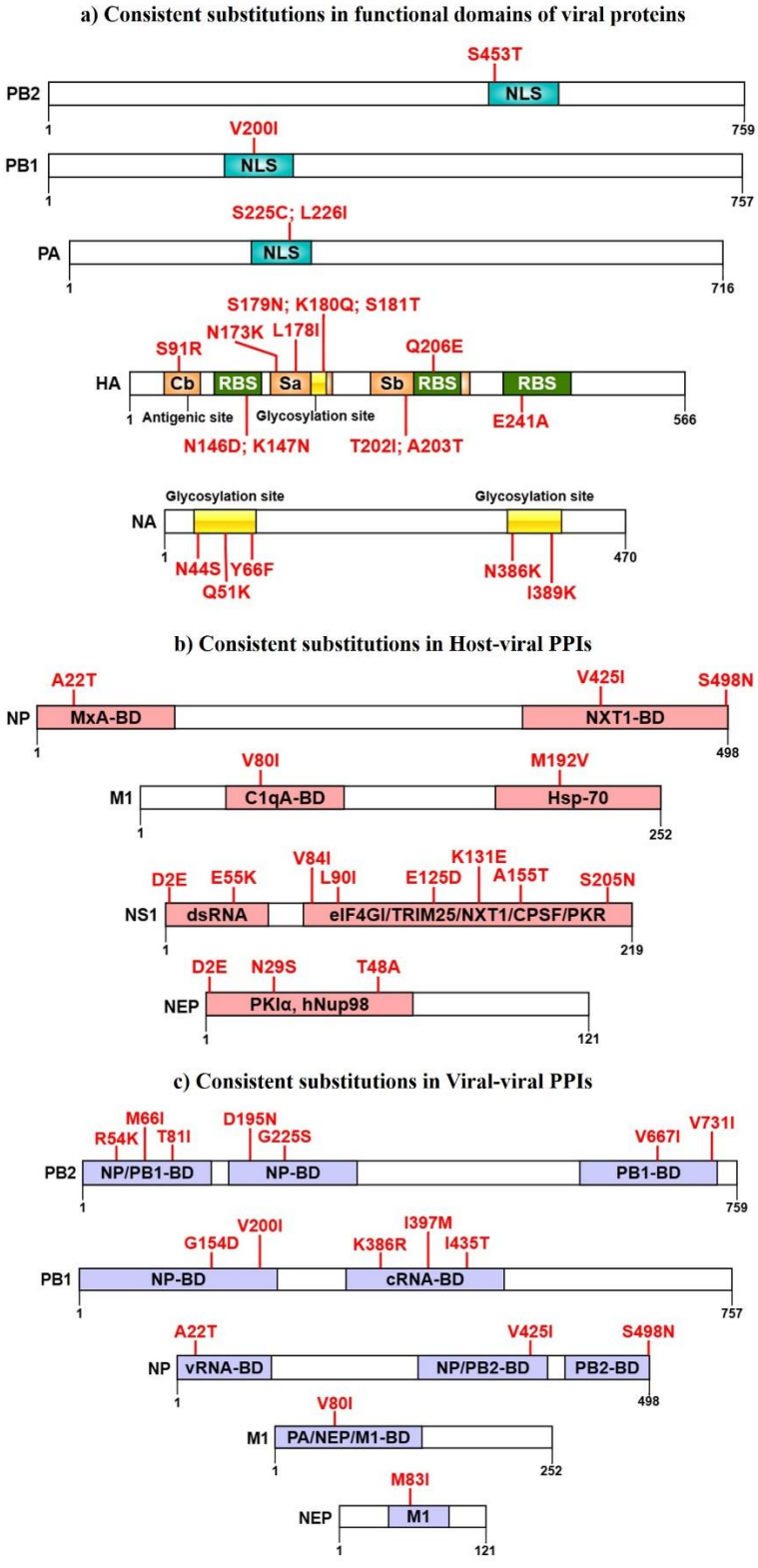


Figure 5: Consortium of “consistent substitutions” on viral proteins engaged at specific stages of the infection

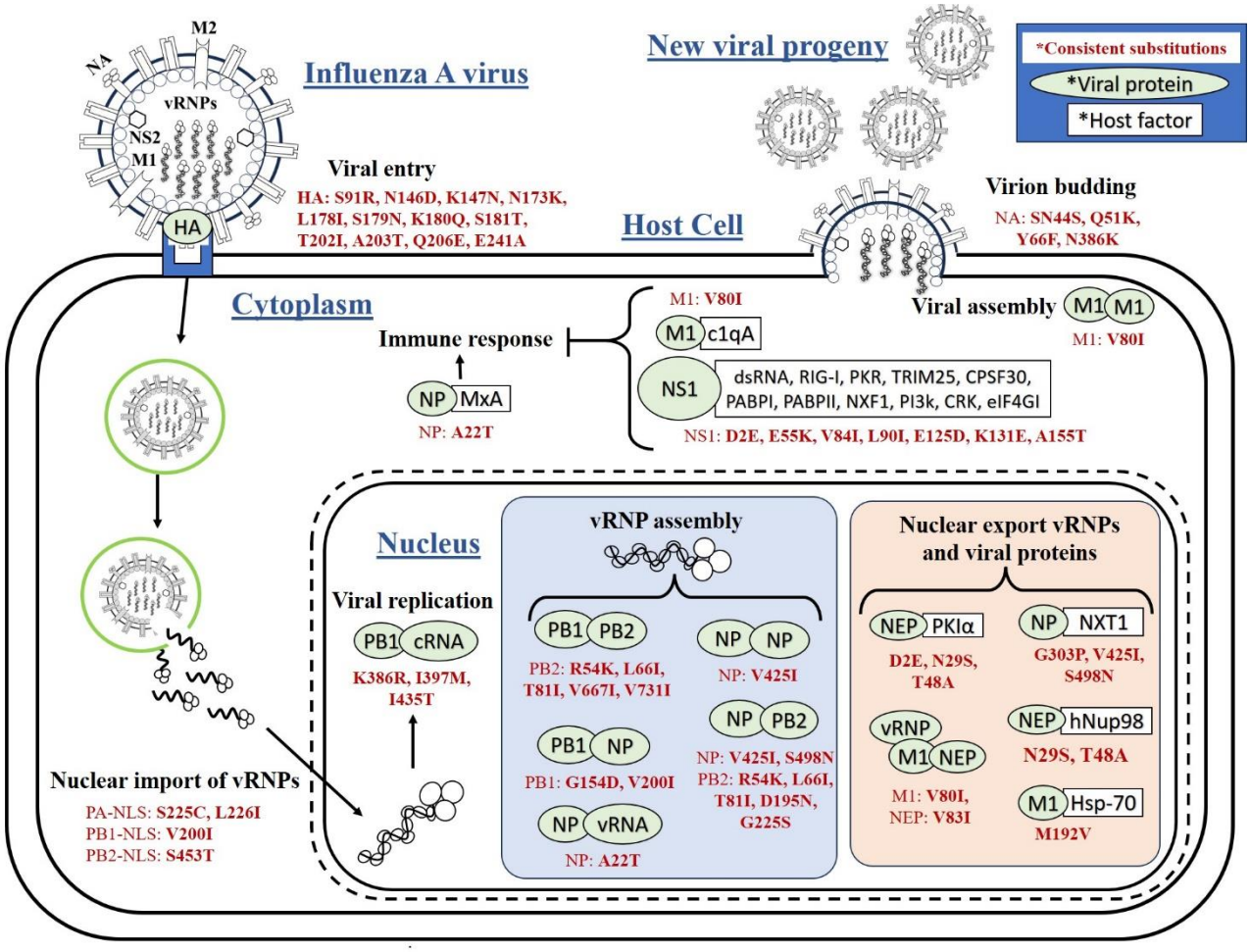


Table 1: List of "consistent substitutions" involved in viral protein functional domains:

A) Nuclear transport signal motifs			
Viral protein	**NLS regions**	**References**	**Consistent substitutions**
PB2	449 to 495, 736 to 739	(8)	S453T
PB1	187 to 211		V200I
PA	124 to 139, 186 to 247		S225C, L226I
B) Receptor binding sites (RBS) on HA protein			
RBS	**Residue position**	**References**	**Consistent substitutions**
130-loop	144 to 154	(2,3)	N146D, K147N
190-helix	204 to 212		Q206E
220-loop	234 to 244		E241A
C) Antigenic sites on HA protein			
Antigenic sites	**Residue positions**	**References**	**Consistent substitutions**

Sa	141-142, 170-174, 176-181	(18)	N173K, L178I, S179N, K180Q, S181T
Sb	201-212		T202I, A203T, Q206E
Cb	87-92		S91R

D) Post-translation modification (Glycosylation)

Viral Protein	Glycosylation sites	References	Consistent Substitutions
HA	179-182	(2,18,25,26,40,41)	S179N, K180Q, S181T
NA	42-45		N44S
NA	50-53		Q51K
NA	63-66		Y66F
NA	386-389		N386K, I389K

Table 2: "Consistent substitutions" involved in changes of glycosylation sites on a) HA and b) NA proteins: "-" indicates the absence of the glycosylation sites.

A)

YEAR	CONSISTENT SUBSTITUTIONS	179-182 (SQSY*)
2017	S179N	NQSY
2018	S181T	NQTY

*Not an experimentally determined glycosylation site

B)

YEAR	CONSISTENT SUBSTITUTIONS	42-45 -	50-53 (NQSV)	63-66 (NQTY)	386-389 (NFSI)
2015	N44S; N386K	NQSQ	NQSV	NQTY	-
2019	Q51K	NQSQ	NKSV	NQTY	-
2021	Y66F	NQSQ	NKSV	NQTF	-

Table 3: List of "consistent substitutions" involved in protein-protein interactions (PPIs): A) Human host-viral protein interactions, B) Viral-viral protein interactions (vPPIs)

A)

Host Factors	*Viral protein*	*Viral protein region involved in PPI*	*Consistent substitutions*	*Cellular and viral functions that might be affected by "consistent substitutions."*	*References*
dsRNA	**NS1**	RNA- Binding domain (1 to 73)	D2E, E55K	Inhibition of OAS/RNase L, Jun N-terminal kinase and RIG-I	(30,43,50)
TRIM25	**NS1**	Effector domain (85 to 219)	L90I, E125D, K131E, A155T, N205S	TRIM25 mediated IFN response, RIG-I CARD ubiquitination, and host interferon system	(20,30,52)
PKR	**NS1**	123-127	E125D	PKR activation blockage	(20,30,43,50)
PABPI and PABPII	**NS1**	Effector domain (85 to 215)	L90I, E125D, K131E, A155T, N205S	PABPI enhances translation of viral mRNA and PABPII blocks the nuclear export of host mRNA.	(20,30,47–49)
NXF1 complex	**NS1**	101 to 155	E125D, K131E, A155T	Blocks nuclear export of host mRNA	(4)
CPSF	**NS1**	Effector domain (85 to 219)	L90I, E125D, K131E, A155T, N205S	viral protein synthesis by recruitment of viral mRNAs to multi-protein translation-initiation complexes	(1,20,51)

PI3K	**NS1**	89-164	L90I, E125D, K131E, A155T	PI3K activation and subsequent enhancement of Akt phosphorylation	(1,30,42,43)
CRK/CRKL	**NS1**	Effector domain (85 to 219)	L90I, E125D, K131E, A155T, N205S	Premature cell death and viral replication	(1,30,44–46)
eIF4GI	**NS1**	81-113	V84I, L90I	Enhances translation of viral mRNA.	(1,20,30,51)
Complement c1qA	**M1**	N-terminal domain (1 to 90)	V80I	Classical complement pathway	(34)
Hsp-70 host protein	**M1**	102-201	M192V	Nuclear export of vRNPs	(53)
NTF-2-NXT1	**NP**	C-terminal region (161 to 498)	V425I, S498N	Nuclear export of NP	(54)
hNup98	**NEP**	22-53	N29S, T48A	Nuclear export of NEP	(55)
Nuclear MxA protein	**NP**	N-terminal domain (1-119)	A22T	Inhibits the transcription of the viral genome	(35)
PKIα	**NEP**	N-terminal domain (1 to 53)	D2E, N29S, T48A	Nuclear transport of viral proteins	(56)

B)

Viral protein 1	**Interaction region of viral protein 1 (reference)**	**Consistent substitutions**	**Viral protein 2**	**Interaction region of viral protein 2 (reference)**	**Consistent substitutions**
M1	N-terminal domain 33 to 158 (57)	V80I	**M1**	N-terminal domain 33 to 158 (57)	V80I
M1-vRNP	76 to 115 (55)	V80I	**NEP**	C-terminal residues 78 and 81 to 100 (55,58)	M83I
PB1	1 to 139 and 267 to 493 (38)	K386R, I397M, I435T	**cRNA**	Unknown	-
NP	1 to 77 and 91 to 188 (8)	A22T	**vRNA**	Unknown	-
NP	340 to 498 (37,59–61)	V425I, S498N	**PB2**	N-terminal domain 1–269 (37,59–61)	R54K, M66I, T81I, D195N, G225S
NP	Unknown	-	**PB1**	1 to 269 and 580 to 638 (59)	G154D, V200I
NP	180 to 336 and 371 to 465 (8)	V425I	**NP**	180 to 336 and 371 to 465 (8)	V425I
PB2	1 to 124 and 577 to 736 (8,37)	R54K, M66I, T81I, V667I, V731I	**PB1**	C-terminal domain 506 to 659 (8,37)	-

*Abbreviations: Cholesterol recognition amino acid consensus (CRAC), Nuclear transport factor 2 (NTF-2)-like export protein1 (NXT1), Ccr4-Not transcription complex subunit 4 (CNOT4), Transcription initiation factor (eIF4GI), Chromodomain-helices-DNA-binding protein 3 (CHD3), Nuclear localisation signal (NLS), Nuclear export signal (NES).

References:

1. Zhao M, Wang L, Li S. Influenza A virus-host protein interactions control viral pathogenesis. Int J Mol Sci. 2017;18(8):1–15.
2. De Graaf M, Fouchier RAM. Role of receptor binding specificity in influenza A virus transmission and pathogenesis. EMBO J. 2014;33(8):823–41.
3. Tharakaraman K, Sasisekharan R. Influenza surveillance: 2014-2015 H1N1 "swine"-derived influenza viruses from India. Cell Host Microbe [Internet]. 2015;17(3):279–82. Available from: http://dx.doi.org/10.1016/j.chom.2015.02.019

4. Zhang Y, Xu Z, Cao Y. Host – Virus Interaction : How Host Cells Defend. 2020;1–23.
5. Hashem AM, Azhar EI, Shalhoub S, Turki ·, Abujamel S, Othman NA, et al. Genetic characterization and diversity of circulating influenza A/ H1N1pdm09 viruses isolated in Jeddah, Saudi Arabia between 2014 and 2015. Arch Virol [Internet]. 2018;163:1219–30. Available from: https://doi.org/10.1007/s00705-018-3732-y
6. Brown EG. Influenza virus genetics. Biomed Pharmacother. 2000;54(4):196–209.
7. Samji T. Influenza A: Understanding the viral life cycle. Yale J Biol Med. 2009;82(4):153–9.
8. Boulo S, Akarsu H, Ruigrok RWH, Baudin F. Nuclear traffic of influenza virus proteins and ribonucleoprotein complexes. Virus Res. 2007;124(1–2):12–21.
9. Noda T, Kawaoka Y. Structure of influenza virus ribonucleoprotein complexes and their packaging into virions. Rev Med Virol. 2010;20(6):380–91.
10. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A. 1992;89(22):10915–9.
11. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547–9.
12. Yang X, Lian X, Fu C, Wuchty S, Yang S, Zhang Z. HVIDB: A comprehensive database for human-virus protein-protein interactions. Brief Bioinform. 2021;22(2):832–44.
13. Patil A, Anhlan D, Ferrando V, Mecate-Zambrano A, Mellmann A, Wixler V, et al. Phosphorylation of Influenza A Virus NS1 at Serine 205 Mediates Its Viral Polymerase-Enhancing Function. J Virol. 2021;95(6).
14. Nogales A, Piepenbrink MS, Wang J, Ortega S, Basu M, Fucile CF, et al. A highly potent and broadly neutralizing H1 influenza-specific human monoclonal antibody. Sci Rep [Internet]. 2018;8(1):1–15. Available from: http://dx.doi.org/10.1038/s41598-018-22307-8
15. Zhang L, Wang J, Muñoz-Moreno R, Kim M, Sakthivel R, Mo W, et al. Influenza Virus NS1 Protein-RNA Interactome Reveals Intron Targeting. J Virol. 2018;92(24):1–16.
16. Liu Y, Wang Y, Liu B, Cong X, Ji Y, Guo X, et al. Phylogenetic analysis and clinical characteristics of the co-occurring mutations in HA and NA genes of influenza A(H1N1)pdm09 viruses during 2015–2017 in Beijing, China. Virol J [Internet]. 2020;17(1):3–11. Available from: https://doi.org/10.1186/s12985-020-01446-3
17. Xu R, Zhu X, McBride R, Nycholat CM, Yu W, Paulson JC, et al. Functional Balance of the Hemagglutinin and Neuraminidase Activities Accompanies the Emergence of the 2009 H1N1 Influenza Pandemic. J Virol. 2012;86(17):9221–32.
18. Igarashi M, Ito K, Yoshida R, Tomabechi D, Kida H, Takada A. Predicting the antigenic structure of the pandemic (H1N1) 2009 influenza virus hemagglutinin. PLoS One. 2010;5(1):1–7.
19. Hara K, Shiota M, Kido H, Watanabe K, Nagata K, Toyoda T. Inhibition of the protease

activity of influenza virus RNA polymerase PA subunit by viral matrix protein. Microbiol Immunol. 2003;47(7):521–6.

20. Hale BG, Randall RE, Ortin J, Jackson D. The multifunctional NS1 protein of influenza A viruses. J Gen Virol. 2008 Oct;89(10):2359–76.
21. Lin YC, Jeng KS, Lai MMC. CNOT4-mediated ubiquitination of influenza a virus nucleoprotein promotes viral RNA replication. MBio. 2017;8(3).
22. Reinhardt J, Wolff T. The influenza A virus M1 protein interacts with the cellular receptor of activated C kinase (RACK) 1 and can be phosphorylated by protein kinase C. Vet Microbiol. 2000;74(1–2):87–100.
23. Manzoor R, Igarashi M, Takada A. Influenza A virus M2 protein: Roles from ingress to egress. Int J Mol Sci. 2017;18(12):1–16.
24. Gao S, Wu J, Liu R-Y, Li J, Song L, Teng Y, et al. Interaction of NS2 with AIMP2 Facilitates the Switch from Ubiquitination to SUMOylation of M1 in Influenza A Virus-Infected Cells. J Virol. 2015;89(1):300–11.
25. Bao D, Xue R, Zhang M, Lu C, Ma T, Ren C, et al. N-Linked Glycosylation Plays an Important Role in Budding of Neuraminidase Protein and Virulence of Influenza Viruses. J Virol. 2021;95(3):1–11.
26. Altman MO, Angel M, Košík I, Trovão NS, Zost SJ, Gibbs JS, et al. Human influenza a virus hemagglutinin glycan evolution follows a temporal pattern to a glycan limit. MBio. 2019;10(2).
27. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomput. 2002;322:310–22.
28. Li W, Godzik A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658–9.
29. Schjoldager KT, Narimatsu Y, Joshi HJ, Clausen H. Global view of human protein glycosylation pathways and functions. Nat Rev Mol Cell Biol [Internet]. 2020;21(12):729–49. Available from: http://dx.doi.org/10.1038/s41580-020-00294-x
30. Lubna S, Chinta S, Burra P, Vedantham K, Ray S, Bandyopadhyay D. New substitutions on NS1 protein from influenza A (H1N1) virus: Bioinformatics analyses of Indian strains isolated from 2009 to 2020. Heal Sci Reports. 2022;5(3).
31. Hussein HA, Geneix C, Cauvin C, Marc D, Flatters D, Camproux AC. Molecular dynamics simulations of influenza a virus NS1 reveal a remarkably stable RNA-binding domain harboring promising druggable pockets. Viruses. 2020;12(5):1–17.
32. Naceri S, Marc D, Camproux AC, Flatters D. Influenza A Virus NS1 Protein Structural Flexibility Analysis According to Its Structural Polymorphism Using Computational Approaches. Int J Mol Sci. 2022;23(3).

33. Cunha AES, Loureiro RJS, Simões CJV, Brito RMM. Unveiling New Druggable Pockets in Influenza Non-Structural Protein 1: NS1–Host Interactions as Anti-viral Targets for Flu. Int J Mol Sci. 2023;24(3).
34. Zhang J, Li G, Liu X, Wang Z, Liu W, Ye X. Influenza A virus M1 blocks the classical complement pathway through interacting with C1qA. J Gen Virol. 2009;90(11):2751–8.
35. Turan K, Mibayashi M, Sugiyama K, Saito S, Numajiri A, Nagata K. Nuclear MxA proteins form a complex with influenza virus NP and inhibit the transcription of the engineered influenza virus genome. Nucleic Acids Res. 2004;32(2):643–52.
36. Chauhan RP, Gordon ML. An overview of influenza A virus genes, protein functions, and replication cycle highlighting important updates. Virus Genes [Internet]. 2022;58(4):255–69. Available from: https://doi.org/10.1007/s11262-022-01904-w
37. Ng AKL, Chan WH, Choi ST, Lam MKH, Lau KF, Chan PKS, et al. Influenza polymerase activity correlates with the strength of interaction between nucleoprotein and PB2 through the host-specific residue K/E627. PLoS One. 2012;7(5).
38. González S, Ortín J. Distinct regions of influenza virus PB1 polymerase subunit recognize vRNA and cRNA templates. EMBO J. 1999;18(13):3767–75.
39. Liu W, Xie Y, Ma J, Luo X, Nie P, Zuo Z, et al. IBS: An illustrator for the presentation and visualization of biological sequences. Bioinformatics. 2015;31(20):3359–61.
40. McKimm-Breschkin JL, Sahasrabudhe A, Blick TJ, McDonald M, Colman PM, Hart GJ, et al. Mutations in a Conserved Residue in the Influenza Virus Neuraminidase Active Site Decreases Sensitivity to Neu5Ac2en-Derived Inhibitors. J Virol. 1998;72(3):2456–62.
41. Kelsey C. Martin Mhatre V. Ho J-AL, Ruben Martin and Stephen L. Buchwald, Mhatre V. Ho and Kelsey C. Martin J-AL, Craik C, Manuscript A, Kantrowitz. 基因的改变NIH Public Access. Bone [Internet]. 2008;23(1):1–7. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf
42. Lopes AM, Domingues P, Zell R, Hale BG. Structure-Guided Functional Annotation of the Influenza A Virus NS1 Protein Reveals Dynamic Evolution of the p85β-Binding Site during Circulation in Humans. J Virol. 2017;91(21):1–16.
43. Hale BG, Kerry PS, Jackson D, Precious BL, Gray A, Killip MJ, et al. Structural insights into phosphoinositide 3-kinase activation by the influenza A virus NS1 protein. 2009;
44. Dubrow A, Lin S, Savage N, Shen Q, Cho JH. Molecular basis of the ternary interaction between NS1 of the 1918 influenza A virus, PI3K, and CRK. Viruses. 2020;12(3):1–10.
45. Hale BG, Randall RE, Ortı J, Jackson D, Jackson D. The multifunctional NS1 protein of influenza A viruses. 2008;2359–76.
46. Shen Q, Zeng D, Zhao B, Bhatt VS, Li P, Cho JH. The Molecular Mechanisms Underlying

the Hijack of Host Proteins by the 1918 Spanish Influenza Virus. ACS Chem Biol. 2017;12(5):1199–203.

47. Jones DT, Taylor WR, Thornton JM, Shapira SD, Gat-viks I, Shum BO V, et al. Appearance of L90I and N205S Mutations in Effector Domain of NS1 Gene of pdm (09) H1N1 Virus from India during 2009-2013. J Virol [Internet]. 2009;8(3):1–10. Available from: http://dx.doi.org/10.1038/nature08182
48. Li W, Wang G, Zhang H, Shen Y, Dai J, Wu L, et al. Inability of NS1 protein from an H5N1 influenza virus to activate PI3K/Akt signaling pathway correlates to the enhanced virus replication upon PI3K inhibition. Vet Res. 2012;43(1):1–12.
49. Tripathi S, Pohl MO, Zhou Y, Rodriguez-frandsen A, Wang G, Stein DA, et al. HHS Public Access. 2016;18(6):723–35.
50. Cheng A, Cheng A, Wong SM, Yuan YA. Structural basis for dsRNA recognition by NS1 protein of influenza A virus. 2009;187–95.
51. Thulasi Raman SN, Zhou Y. Networks of host factors that interact with NS1 protein of influenza a virus. Front Microbiol. 2016;7(MAY):1–15.
52. Koliopoulos MG, Lethier M, Van Der Veen AG, Haubrich K, Hennig J, Kowalinski E, et al. Molecular mechanism of influenza A NS1-mediated TRIM25 recognition and inhibition. Nat Commun [Internet]. 2018;9(1). Available from: http://dx.doi.org/10.1038/s41467-018-04214-8
53. Watanabe K, Fuse T, Asano I, Tsukahara F, Maru Y, Nagata K, et al. Identification of Hsc70 as an influenza virus matrix protein (M1) binding factor involved in the virus life cycle. FEBS Lett. 2006;580(24):5785–90.
54. Chutiwitoonchai N, Aida Y. NXT1, a novel influenza A NP binding protein, promotes the nuclear export of NP via a CRM1-dependent pathway. Viruses. 2016;8(8):1–15.
55. Shimizu T, Takizawa N, Watanabe K, Nagata K, Kobayashi N. Crucial role of the influenza virus NS2 (NEP) C-terminal domain in M1 binding and nuclear export of vRNP. FEBS Lett [Internet]. 2011;585(1):41–6. Available from: http://dx.doi.org/10.1016/j.febslet.2010.11.017
56. O'Neill RE, Palese P. NPI-1, the human homolog of SRP-1, Interacts with influenza virus nucleoprotein. Virology. 1995;206(1):116–25.
57. Safo MK, Musayev FN, Mosier PD, Zhou Q, Xie H, Desai UR. Crystal structures of influenza a virus matrix protein M1: Variations on a theme. PLoS One. 2014;9(10).
58. Akarsu H, Burmeister WP, Petosa C, Petit I, Müller CW, Ruigrok RWH, et al. Crystal structure of the M1 protein-binding domain of the influenza A virus nuclear export protein (NEP/NS2). EMBO J. 2003;22(18):4646–55.
59. Poole E, Elton D, Medcalf L, Digard P. Functional domains of the influenza A virus PB2 protein: Identification of NP- and PB1-binding sites. Virology. 2004;321(1):120–33.

60. Gabriel G, Herwig A, Klenk HD. Interaction of polymerase subunit PB2 and NP with importin α1 is a determinant of host range of influenza A virus. PLoS Pathog. 2008;4(2).
61. Szeto WC, Hsia HP, Tang YS, Shaw PC. Interaction between influenza A virus nucleoprotein and PB2 cap-binding domain is mediated by RNA. PLoS One [Internet]. 2020;15(9 September):1–11. Available from: http://dx.doi.org/10.1371/journal.pone.0239899