Virology 535 (2019) 297–307

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology

To hit or not to hit: Large-scale sequence analysis and structure characterization of influenza A NS1 unlocks new antiviral target potential

João M. Trigueiro-Louro[a,b,**], Vanessa Correia[b], Luís A. Santos[b], Rita C. Guedes[c], Rui M.M. Brito[d], Helena Rebelo-de-Andrade[a,b,*]

[a] Host-Pathogen Interaction Unit, Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, Universidade de Lisboa, Av. Professor Gama Pinto, 1649-003, Lisbon, Portugal

[b] Antiviral Resistance Lab, Research & Development Unit, Infectious Diseases Department, Instituto Nacional de Saúde Doutor Ricardo Jorge, IP, Av. Padre Cruz, 1649-016, Lisbon, Portugal

[c] Medicinal Chemistry Unit, Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, Universidade de Lisboa, Av. Professor Gama Pinto, 1649-003, Lisbon, Portugal

[d] Chemistry Department and Coimbra Chemistry Centre, Faculty of Science and Technology, University of Coimbra, 3004-535, Coimbra, Portugal

ARTICLE INFO

Keywords:
Anti-influenza strategy
Consensus druggable pocket
Conservation
Druggability
Influenza
Non-structural protein 1

ABSTRACT

Influenza NS1 protein is among the most promising novel druggable anti-influenza target, based on its structure; multiple interactions; and global function in influenza replication and pathogenesis. Notwithstanding, drug development guidance based on NS1 structural biology is lacking. Here, we design a promising strategy directed to highly conserved druggable regions as a result of an exhaustive large-scale sequence analysis and structure characterization of NS1 protein across human-infecting influenza A subtypes, over the past 100 years. We have identified 3 druggable pockets and 8 new potential hot spot residues in the NS1 protein, not described before, additionally to other 16 sites previously identified, which represent attractive targets for pharmacological modulation. This study provides a rationale towards structure-function studies of NS1 druggable sites, which have the potential to accelerate the NS1 target validation. This research also contributes to a deeper comprehension and insight into the evolutionary dynamics of influenza A NS1 protein.

1. Introduction

Annual epidemics of human influenza A virus (IAV) cause a significant economic burden, morbidity and mortality, worldwide (Cox, 1999; Krammer et al., 2018). Influenza antiviral drugs are the only specific and directed available treatment, particularly important in case of a newly emerging influenza virus or in specific groups (pregnant women, immunocompromised hosts, individuals with chronic illness, children and the elderly patients) (Ghebrehewet et al., 2016). An effective antiviral therapy is determinant to improve the patient prognosis, to reduce morbidity, mortality and to minimize the economic burden in the health care system (Klepser, 2014; Putri et al., 2018).

Considering the emergence of resistance to the first class of influenza antivirals, neuraminidase inhibitors are the anti-influenza drugs of choice (Hurt et al., 2016). Current influenza antivirals present known limitations in scope, effectiveness and emergence of resistance strains (Takashita et al., 2015). Since the virus evolve at a high rate, the clinical management of influenza is constantly threatened by the lack of therapeutic choices, which highlights the need for new antiviral drug discovery and development that explore a wider range of highly conserved targets (Penttinen and Catchpole, 2016).

The mindset of attrition during drug discovery and development is repeatedly in debate, since only about 4% of drug development projects lead to licensed drugs (Mohs and Greig, 2017; Schuhmacher et al., 2016). Undruggable targets are responsible for up to 60% of failure in this process (Mohs and Greig, 2017; Schuhmacher et al., 2016). In this context, a critical step in any new early drug discovery approach abide to predict the druggability of the target longing to explore. Furthermore, established drug targets are correlated to evolutionary conservation since highly conserved regions are often biologically

* Corresponding author. Host-Pathogen Interaction Unit, Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, Universidade de Lisboa, Av. Professor Gama Pinto, 1649-003, Lisbon, Portugal

** Corresponding author. Antiviral Resistance Lab, Research & Development Unit, Infectious Diseases Department, Instituto Nacional de Saúde Doutor Ricardo Jorge, IP, Av. Padre Cruz, 1649-016, Lisbon, Portugal

E-mail addresses: joao.louro@insa.min-saude.pt (J.M. Trigueiro-Louro), H.Rebelo.Andrade@insa.min-saude.pt (H. Rebelo-de-Andrade).

https://doi.org/10.1016/j.virol.2019.04.009
Received 19 February 2019; Received in revised form 22 April 2019; Accepted 23 April 2019
Available online 27 April 2019
0042-6822/

Abbreviations

aa	amino acid
CDP	Consensus druggable pocket
CTT	C-terminal "tail"
DGSS	DoGSiteScorer
ED	effector domain
H1N1pdm09	pandemic H1N1 2009
H1N1sea	seasonal H1N1
IAV	influenza A virus
LR	linker region
LV	length variant
NS1	non-structural protein 1
PDB	Protein Data Bank
PDS	PockDrug-Server
RBD	RNA-Binding Domain
SF	SiteFinder from MOE

important for virus replication and display essential structural or functional roles during infection (Lv et al., 2015). In this regard, the identification of binding pockets within conserved regions should result in more robust antiviral targets since they are less prone to mutations and evade the host immune response.

Structure-based drug design approaches provide an efficient way to direct cost-intensive experimental efforts and have already proven fundamental in the discovery and development of several drugs, including the anti-influenza zanamivir and oseltamivir (Kirchmair et al., 2011; Mallipeddi et al., 2014).

The influenza non-structural (NS) gene corresponds to the viral RNA segment 8 which encodes both the NS1 and the viral nuclear export protein (NEP) (Engel, 2013). Structurally, the non-structural protein 1 (NS1) is comprised by two functional domains bridged by a short inter-domain linker region (LR): N-terminal RNA-Binding Domain (RBD) (amino acids 1–73); and C-terminal effector domain (ED) (amino acids 86–202) (Hale, 2014; Hale et al., 2008). The remaining residues of NS1 form a disordered C-terminal "tail" (CTT) (Marc, 2016). The protein is able to form homodimers (RBD and ED) and multimers (only for ED) (Hale, 2014; Hale et al., 2008; Krug, 2015).

NS1 emerges as a potential target for antiviral drug discovery based on its structure; interaction with host proteins and nucleic acids; and global function in influenza replication (Hale et al., 2008). NS1 protein is among the most promising novel anti-influenza targets. It is exclusively and highly expressed in virus-infected cells (both in cytoplasm and nucleus), being absent from the virion particle (Abdelwhab et al., 2013; Marc, 2014). It mediates crucial functions in virus replication, propagation and pathogenesis by way of the interplay with multiple protein-RNA and protein-protein interactions (Marc, 2014).

The N-terminal RBD forms an obligate six-helical homodimer (the dimer interface consists of two antiparallel α-helixes) which binds to several RNAs (Kerry et al., 2011).

The effector domain can also dimerize and forms distinct quaternary conformations: It is established that at the early stages of infection the NS1-ED co-exist in a monomeric and dimeric states (Carrillo et al., 2014). The monomer promotes the interaction of NS1 with cellular proteins, such as p85β subunit of phosphatidylinositol 3-kinase (PI3K) (Carrillo et al., 2014; Hale et al., 2010a). At later stages of infection, the homodimeric state of NS1 may be the predominant form. Distinct ED-ED interaction types and dimer interfaces have been observed in the influenza NS1 protein, so the precise fashion in which the ED monomers can dimerize is unclear (Bornholdt and Prasad, 2008; Carrillo et al., 2014; Hale et al., 2008; Kerry et al., 2011; Xia and Robertus, 2010). It has been suggested that the helix-helix dimer is the predominant conformation of NS1 in the homodimeric state, while the strand-strand dimer may exist as a minor form (Aramini et al., 2014; Carrillo et al., 2014; Hale et al., 2008; Kerry et al., 2011; Xia and Robertus, 2010). Additionally, the NS1-ED, but not the NS1-RBD, multimerizes in order to manage the pleiotropic function of NS1 (Hale, 2014; Wang et al., 1999).

Table 1
Length variations and C-terminal sequence of the influenza A virus NS1 protein.

		NS1 Length Variations (LV) Types													
	IAV Subtype	202	212	215	217	219	220	224	225	227	228	230	237	Total	C-terminus sequence: only for LV types with a prevalence >1%
IAV previously established in humans	prior H1N1sea	2			4	**277**		1			1	1314	**105**	1704 (6,002%)	219 aa: ^{212}PSLPPEQK^{219} 230 aa: ^{212}PPFT(T)/(P)TQK^{219}; ^{227}RSEV^{230} (pre-2009) 237 aa: ^{212}PPLTPKQK^{219}; ^{227}RSEVRRNKMAD^{237} (pre-1989)
	H1N1pdm09	1			4	11474						11		11490 (40,469%)	219 aa: ^{212}PSLPPEQK^{219} (post-2009)
	H1N2					**14**	**1**					31		46 (0,162%)	219 aa: ^{212}PSLPPEQK^{219} (post-2007) 230 aa: ^{227}RSKV^{230} (pre-2004)
	H2N2				1		**14**					**2**	86	103 (0,363%)	220 aa: ^{213}PLTPKQKR^{220} (1967-1968) 237 aa: ^{227}RSKVRR(D)/(N)KMAD^{237} (1957-1968)
	H3N2				105	**174**	19			3	1	13713	**208**	14223 (50,095%)	219 aa: ^{212}PSLPPEQK^{219} (*major*) or ^{212}PPLPPEQK^{219} (*minor*) (mostly post-2009) 230 aa: ^{212}PPLTPKQK^{219}; ^{227}RSKV^{230} (1970-2012/2013); ^{227}RSEV^{230} (post-2011/2012) 237 aa: ^{212}PPLTPKQK^{219}; ^{227}RSKVRRDKMAD^{237} (mostly pre-1990)
Avian-to-human transmission	H5N1		1	7					277			**12**		297 (1,046%)	225 aa: ^{212}LPLPPNQK^{219}; ^{222}ESEV^{225} (*major*) or ^{222}ESKV^{225} (*minor*) [contains a 5-aa deletion - 80-84 aa] (post-2000) 230 aa: ^{212}SPLPPKQK^{219}; ^{227}EPEV^{230} (pre-2000)
	H5N6				12				**6**					18 (0,063%)	217 aa: ^{212}PPLSPK^{217} or ^{212}SPFSTK^{217} (post-2015) 225 aa: ^{222}ESEV^{225} [contains a 5-aa deletion - 83-87 aa] (post-2014)
	H7N7											4		4 (0,014%)	230 aa: ^{227}ESEV^{230}
	H7N9				490								2	492 (1,733%)	217 aa: ^{212}SPLSTK^{217} (*major*) or ^{212}PPLSPK^{217} (*minor*) (post-2013)
	H9N2				9							**6**		15 (0,053%)	217 aa: ^{212}SPLSTK^{217} (*major*) or ^{212}PPLSPK^{217} (*minor*) (post-2014) 230 aa: ^{227}EPEV^{230} (pre-2011)
	Total	**3** (0,011%)	**1** (0,004%)	**7** (0,025%)	**625** (2,201%)	**11939** (42,051%)	**34** (0,120%)	**1** (0,004%)	**283** (0,997%)	**3** (0,011%)	**2** (0,007%)	**15093** (53,159%)	**401** (1,412%)	28392	

LV types with a prevalence greater than 1% are marked in bold. The predominant LV type for each IAV subtype is highlighted in blue.
The C-terminus sequences are described in more detail for the most prevalent IAV subtypes (H1N1, H3N2, H5N1 and H7N9) from the aa 212-237, when applied.

The NS1-ED interacts with multiple cellular proteins and host factors including the ubiquitin E3 ligases: tripartite motif-containing 25 (TRIM25) and Riplet; the RNA-dependent protein kinase (PKR); the 30-kDa subunit of the cleavage and polyadenylation specificity factor (CPSF30); the eukaryotic translation initiation factor 4GI (eIF4GI); the poly(A)-binding protein II (PABPII); and the p85β subunit of PI3K, etc. (Kleinpeter et al., 2018; Robb et al., 2010). These interactions will affect several cellular pathways, which will globally contribute to the major function of NS1: evade the host innate immune response predominantly by blocking the type I interferon (IFN); and lead to the inherent infection pathogenicity (Garcia-Sastre, 2017; Kleinpeter et al., 2018; Robb et al., 2010).

The ED has an intrinsic plasticity behaviour which facilitates NS1 pleiotropic functions in a spatial and temporal fashion (Aramini et al., 2014). In this regard, the NS1-ED may either regulate the inhibition of cellular apoptosis (at early stages) or the induction apoptosis (at later stages); and inhibit IFN-inducing proteins or, contrarily, promote the induction of type I IFNs in a fraction of the infected cells (Abdelwhab et al., 2013; Das et al., 2008; Hrincius et al., 2011; Kerry et al., 2011; Marc, 2014). The conformational plasticity of NS1 also contributes to stabilize the RBD. NS1 multifunctionality is facilitated by its own autoregulation which may control the conformational, spatial and temporal distribution of the protein in distinct stages of the infectious cycle (Aramini et al., 2014; Bornholdt and Prasad, 2008; Hale et al., 2009; Hsiang et al., 2012; Kerry et al., 2011). In this context, both previously described interactions (NS1-RBD:RNA, NS1-ED:CPSF30, NS1-ED:p85β) and other unknown NS1:RNA and NS1:protein interactions should be further elucidated and targeted for drug discovery (Engel, 2013; Hale et al., 2010a; Hsu, 2018).

To date, there are no clinical drugs targeting this protein and drug development guidance focused on NS1 structural biology is lacking since the study of anti-NS1 inhibitors is still taking its first steps.

This study aim at identifying putative antiviral target regions within the NS1 influenza protein, mainly focused on undisclosed regions in the NS1 effector domain. In this regard, we intend to identify highly conserved regions or sites located on predicted binding pockets of NS1-ED, across a wide range of human-infecting Influenza A subtypes. Such approach also intends to comprehensively characterize the sequence-to-structure NS1 features in the human-infecting influenza viruses and reveal determinants of protein structure motifs that may, in turn, disclose potential hot spots for drug targeting.

2. Results and discussion

2.1. NS1 length variation studies

A large-scale analysis of the length variation of a total of 28392 NS1 protein sequences from human-infecting influenza A viruses have shown that the protein most frequently contains 230 residues (≈53,2%). However, the precise length diverges among influenza virus subtypes and strains: premature stop codons, or contrarily, the suppression of the most prevailing stop codon (at nucleotide 688–690) produce a protein with a length variation that may range from 202 to 237 residues, as described by other authors (Hale et al., 2008).

Aside from the predominant 230-amino acid (aa) protein, other eleven length variant (LV) types can be found among all the influenza A virus, as depicted in the Table 1. Considering the classification adopted by Dundon et al. (Dundon and Capua, 2009) the following LV types were found, in order of decreasing frequency: 219 aa (42,05%), 217 aa (2,20%), 237 aa (1,41%), 225 aa (≈0,1%); and the 202-aa, 212-aa, 215-aa, 220-aa, 224-aa, 227-aa and 228-aa LV types (accounting for less than 0,1% each on the global length prevalence).

This analysis included a total of 27566 (97,09%) IAV isolates previously established in the human population (H1N1, H1N2, H2N2 and H3N2), and 826 (2,91%) IAV isolates of avian origin (H5N1, H5N6, H7N7, H7N9 and H9N2).

The results showed that the 230-residue protein is the predominant LV type among the prior seasonal H1N1 (H1N1sea), H1N2, H3N2 and H7N7 IAV subtypes. The great majority of variations from the 230-aa protein have been found in a shorter length: the H5N6, H7N9 and H9N2 IAV subtypes mostly exist as a 217-aa protein; the pandemic H1N1 2009 (H1N1pdm09) IAV mostly occurs as a 219-residue protein (99,86%); and the H5N1 IAV NS1 usually contains 225 aa. Only the NS1 from the H2N2 subtype has most often occurred as a lengthy type of 237 residues. The 219-aa protein is limited to the IAV isolates previously established in the human population (H1N1, H1N2 and H3N2), whereas NS1 LV types of 217 aa and 225 aa were only found in IAV of avian origin, namely: the H5N1, H5N6, H7N9 and H9N2.

From the human IAV isolates of avian origin, only the H7N9 IAV subtype (n = 2/825 isolates) has shown to encode a NS1 protein with the 7-aa elongation (237-aa protein), which was previously limited only to the prior H1N1sea, H2N2 and H3N2 IAV subtypes. Human IAV with a NS1 shorter than 217 aa are uncommon (n = 11/28392). The reported length variation can determine functional differences between virus subtypes, as previously reported by other authors (Abdelwhab et al., 2013; Dundon and Capua, 2009; Hale et al., 2010b; Marc, 2016).

In addition to the reported C-terminal deletions, some LV types exhibit a 5-aa deletion within the inter-domain LR, which impacts the protein conservation within these regions. These gaps are confined to the NS1-225 LV type, either in the H5N1 IAV subtype or in the H5N6 IAV subtype. The NS1-225 LV type from the H5N1 IAV contains a 5-aa deletion in the peptide sequence corresponding to the 80–84 aa; and the NS1-225 LV type from the H5N6 IAV encloses a 5-residue deletion in the 83–87 aa sequence.

These results support that the linker region is variable in length. Previous studies have demonstrated that this region is highly flexible, which enables the NS1 protein to adopt distinct conformations regarding its RBD and ED domains. To date, none of the available crystallographic structures of NS1 has been resolved for the linker region (Bornholdt and Prasad, 2008). This poses a major limitation for a structure-based study of this region as a potential antiviral target.

A detailed analysis of the C-terminal 212–219 aa region from human-infecting influenza A virus sequences have revealed that it might be simultaneously IAV subtype- and LV-type-specific. The C-terminal sequence ^{212}PSLPPEQK^{219} is specific of the NS1-219 LV type. The considered amino acid sequence is shared by all H1 IAV subtypes (either H1N1sea, H1N1pdm09 or H1N2); and also by the H3N2 subtype from the same LV group.

Contrarily, the NS1-230 and NS1-237 LV types from the H1N2, H2N2 and H3N2 IAVs display the amino acid sequence ^{212}PPLTPKQK^{219}, together with the NS1-237 LV type from the prior H1N1sea. The NS1-230 LV type from the prior H1N1sea represents an exception since it encodes the sequence ^{212}PPFTTTQK^{219}.

In the same way, the NS1-225 LV type from the H5 IAV subtypes contains the sequence ^{212}LPLPPNQK^{219}; and the NS1-230 LV type from the H5 and H9 IAV subtypes encodes the sequence ^{212}SPLPPKQK^{219}. These results suggest that only the last two amino acids in the considered region (^{218}QK^{219}) may be conserved among all the IAV subtypes.

The ^{212}PPLSPK^{217} and ^{212}SPFSTK^{217} sequences are the most common C-terminus found in the NS1-217 LV type of the H5N6 IAV subtype; and the ^{212}SPLSTK^{217} (*major*) or ^{212}PPLSPK^{217} (*minor*) are the terminal domains for the NS1-217 LV type of the H7N9 and H9N2 IAV subtypes. The NS1-217 LV types represent shorter variants of the predominant NS1-230 LV type; and have only been recently reported in human IAV isolates.

In both NS1-230 and NS1-237 LV types, the C-terminus 227–230 aa region is widely recognized as a potential postsynaptic density 95/disc large/zonula occludens-1 (PDZ) ligand (PL) motif of the X-S/T-X-V type (the X can be any aa) (Jackson et al., 2008). Our dataset endorses the prevailing presence of the ^{227}RSEV^{230} or the ^{227}RSKV^{230} sequences in the NS1 protein from human-infecting influenza A viruses, as

previously reported by other authors (Golebiewski et al., 2011; Jackson et al., 2008; Obenauer et al., 2006). Contrarily, the avian-to-human IAV isolates are characterized by the avian derived NS1 PL sequences of ^{227}ESEV^{230}, ^{227}EPEV^{230} (*major* contribution) or ^{227}ESKV^{230} (*minor* contribution), as previously described by Obenauer et al. This pattern of a major common human C-terminal RSKV domain and a major common avian C-terminal ESEV domain might indicate that this region is under strong selective pressure and it may has been selected during the virus evolution (Obenauer et al., 2006).

The PL region will be addressed for each IAV subtype, as follows. The ^{227}RSKV^{230} domain is shared by the NS1-230 LV types from the H1N2, H2N2 and H3N2 subtypes which have been previously circulated (before 2013) in the human population. Specifically, the H1N2 viruses belonging to the NS1-230 LV type have circulated until 2004 and have been replaced by a shorter NS1-219 LV type, which does not contain the PL motif. The H3N2 subtype has emerged as the "Hong Kong influenza" pandemic in 1968 and exhibited the ^{227}RSKV^{230} sequence (Golebiewski et al., 2011; Jackson et al., 2008); It had been gradually replaced in 2012 by H3N2 IAV containing 230 aa with the ^{227}RSEV^{230} domain or, to a lesser extent, in 2009 by viruses containing a NS1-219 variant lacking the PL domain. The NS1-230 LV types identified in the H5N1 and H9N2 IAV human isolates used to contain the ^{227}EPEV^{230} C-termini. These variants have been replaced by shorter NS1-225 and NS1-217 variants, respectively. The NS1-225 LV type from the H5N1 IAV is characterized by the corresponding ^{222}ESEV^{225} (*major*) or ^{222}ESKV^{225} (*minor*) sequences, along with the NS1-225 LV type from the H5N6 IAV. The ^{222}ESEV^{225} motif is also the PL domain harboured by the H7N7 IAV subtypes.

Finally, the NS1-237 variants are predominantly found among IAV previously established in human population from 1940. The 227–237 region is characterized by the sequence ^{227}RS(E)/(K)VRR(D)/(N)KMAD^{237}. However, our data suggests that the NS1-237 variant is no longer circulating among the human population: since NS1 proteins of prior H1N1sea, H2N2 and H3N2 IAV viruses (isolated since 1989, 1968 and 1990, respectively) lack the 231 to 237 sequence. This information is in concordance with other previous similar studies (Kuo et al., 2016). Recently, two human IAV isolates (2015 and 2017) of the H7N9 subtype have been found with the NS1 7-aa elongation: ^{227}KPEVRRNKMVD^{237} sequence.

The 1918 IAV (A/BrevigMission/1/1918) is a singular variant: It is a virus that contains 230 aa and the C-termini is comprised by the ^{212}PPLPPKQK^{219} and ^{227}KSEV^{230} motifs (Golebiewski et al., 2011), which differ from all the other described NS1 C-terminal domains.

Taken together, these results corroborate that not only NS1 length diverges among influenza virus subtypes and LV types (attributable to variations on the LR and CTT); but also the C-terminal 212–237 aa sequence is highly variable and simultaneously IAV subtype- and LV type-specific.

Viruses that circulate in the human population containing a NS1-237 variant are uncommon since the year of 1990. In this way, designing an antiviral strategy targeting this region is not suitable for the current human circulating IAV. Additionally, since the emergence of the H1N1pdm09 variant in 2009 that the NS1 from both H1 influenza subtypes predominantly exists as a 219-aa protein; although H3N2 IAV continue to circulate with a NS1 protein containing 230 aa. In this way, targeting the C-terminal 220–237 aa region would be only reasonable for a strategy directed for the H3N2 IAV subtype and a few minor variants currently found in some IAV of avian origin. This represents a great limitation, considering that such strategy would only covered a 53,159% of the viruses circulating in the human population. Previous studies have demonstrated that extending the length of the NS1-219 LV type (H1N1pdm09 IAV) to 230 aa does not appear to impact the viral replication efficiency (Hale et al., 2010b). Hence, this region does not satisfy an important criterion for a proper anti-influenza target and we recognize that targeting the C-terminal 220–230 region presents lower antiviral target potential.

NS1 have been found with length variations shorter than the 219-aa LV type, namely the NS1 LV types of 217, 215, 212 and 202 aa which represent less than 2,25% of the total viruses circulating in the human population for the considered time period. Although the prevalence of such short NS1 variants is low, the H7N9 (along with H5N6 and H9N2) IAV subtypes mostly exist as a 217-aa protein. The avian H7N9 IAV subtype was first detected in humans in March 2013 and poses a current serious threat to public health: it causes severe human illness and is

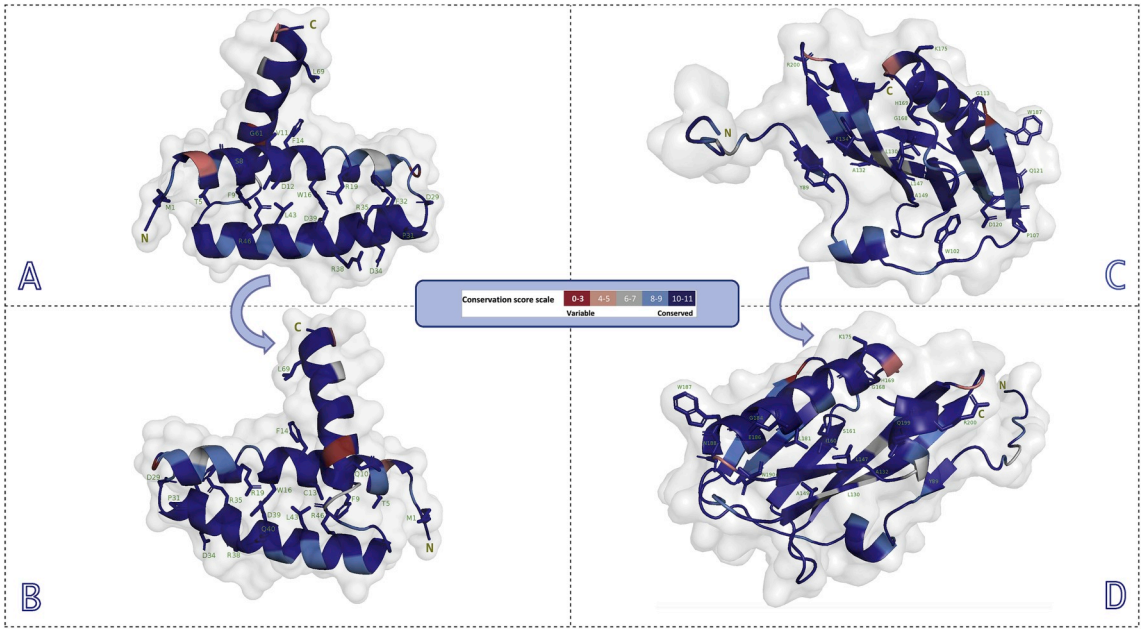


Fig. 1. Three-dimensional sequence conservation of influenza AH1N1pdm09 NS1 RBD (A and B) and ED (C and D) (PDB ID: 3M5R). The conservation scores are mapped onto the protein structures. The N- and C-terminal are indicated. The figure is coloured based on conservation scores: the highest conservation positions are highlighted in shades of blue and the red regions indicate low conservation according to the conservation score scale. Conserved residues with a percent identity score above 99,95% are labelled. Each protein domain is shown in two orientations (180° rotation). The figures were produced with PyMOL Molecular Graphics System (Schrodinger, 2015).

associated with a high mortality rate (Kleinpeter et al., 2018). In this context, a designed drug directed to the C-terminal 218–237 aa region would not be effective against these viruses.

Finally, human IAV containing a NS1 shorter than 217 aa are uncommon, accounting for less than 0,05% on the global length prevalence of the current human circulating IAV. However, an anti-influenza strategy targeting binding sites beyond the 202-aa site should be cautious, considering there are IAV previously established in the human host with a length as short as 202 aa.

Furthermore, the crystallographic structure of CTT, as occurs for the LR, has not been solved, which limits the performance of druggability studies for this region.

Highly variable regions (in length or aa composition) or sites within the NS1 protein may suggest that the integrating residues at that location do not potentially have an important/particular role in protein structure or overall function. Contrarily, other authors have been proposed that some of these sites may be associated with determining host range (as it has been hypothesised for the PDZ domain), NS1 subcellular localization or virus pathogenicity in some IAV subtypes (usually avian IAV); although such characteristics are not universal for all influenza A viruses (Abdelwhab et al., 2013; Dundon and Capua, 2009; Jackson et al., 2008).

With respect to our main goal of identifying highly conserved regions or sites located on predicted binding pockets; and considering that the variable LR and CTT regions are amenable for major modifications (either in length or aa composition), we suggest that neither one should be targeted for anti-influenza A drug discovery.

A good antiviral target should satisfy the important criteria of being predominantly conserved (for instance, sites under low selective pressure) and also play an important role in influenza replication or pathogenesis (Engel, 2013). In this way, our efforts regarding the conservation and druggability studies will be mainly directed to the RBD and the ED regions excluding the variable C-terminal tail (> 202-aa region) in order to fully cover the common regions of the current IAV circulating in the human population.

2.2. Conservation studies

The conservation scores given by the Jalview AACons Web server were calculated for each amino acid position in the NS1 protein. Overall, the human NS1 protein is highly conserved: 66,67% of the residues are highly conserved (conservation score ≥ 10); 87,34% have a conservation score ≥7; and only 9,70% of the residues are highly variable with a score of 0–4. All highly conserved residues (grades 10–11), conserved residues (grades 7–9) and variable residues (grades 1–4) were identified and are depicted in Tables S–1 (supplementary data), along with the identity percentage classification for each residue. The conservation scores were mapped onto the NS1 structure (H1N1pdm09 subtype, PDB entry 3M5R) as illustrated in Fig. 1. High level of conservation was found in several domains of the RBD and the ED, while most of the variable sites were found in the C-terminal tail. Two residues, namely: G61, G158 have shown no variation among the total number of sequences studied. Conversely, variable residues (conservation score between 0 and 4) were found at positions G26, H59 and G71 in the RBD; at position E112 in the ED; and at positions G211 and in the R220-D237 region belonging to the C-terminal tail.

Additional to the study of individual sites, we have identified conserved clusters (of ≥10 continuous residues) on the NS1-RBD and –ED, which are depicted in the Figs. 2 and 3. The average percent identity score of conserved segments was calculated based on the amino acid frequency information at each particular position. The linker-region and the C-terminal tail were excluded from this analysis taking into consideration the sequence deletions and truncations reported in the NS1 length variation study.

The N-terminal NS1-RBD contains two main conserved domains. The first conserved domain, RBD.1, (17-aa length) is comprised by the residues T5-to-Q21; and has a percent identity score of 94,28%. The positions 1, 5–6, 9–16 and 19 exhibit a percent identity score above 99,9%. This region is located on the helix α1 which mostly lies in the buried part of the de RBD dimer. The second conserved cluster, RBD.2, (28-aa length) is formed by the residues L27-to-I54; with a percent identity score of 94,05%. The positions 29, 31–32, 34–35, 37–40, 43, 46 and 50 exhibit a percent identity score above 99,9%. This region is located at the end of the helix α1 and continues to the helix α2 where it forms a major groove along with the antiparallel helix α2'. An additional conserved area of 7 residues (not marked in the image) is composed by the residues V60-to-E66; with a percent identity score of 99,00%. The positions 61 and 69 have a percent identity score above 99,9%. This region is located at the helix α3 which also forms a small groove along with the antiparallel helix α3' from the equivalent monomeric unit.

The NS1-ED contains four conserved clusters along with other four additional smaller conserved areas of 5–10-aa length. The first conserved domain, ED.1, (24-aa length) is formed by the aa R88-to-V111; and has a percent identity score of 86,89%. The positions 89, 97, 100, 102, 105, 107 and 110 exhibit a percent identity score above 99,9%. This region corresponds to the strand β1, the helix α4 and the strand β2 structures of the ED; and is surface exposed.

The second domain, ED.2, (16-aa length) with a percent identity score of 89,37% is formed by residues G113-to-I128. Only the positions 113, 120–121 and 126 exhibit a percent identity score above 99,9%. This region encompasses the strand β3 (which is mostly found buried inside the protein) along with the initial part of the antiparallel strand β4' which is surface exposed. This region is followed by two short conserved areas with highly conserved A132-F138 (7-aa length) and R140-L144 (7-aa length) residues, respectively.

The third conserved cluster, ED.3, (26-aa length) represents the larger cluster in the NS1-ED; with a percent identity score of 97,89%. It is comprised by the aa L146-to-T170. The sites 147-to-149, 151, 158-to-161, 163, 165 and 168-to-170 exhibit a percent identity score above 99,9%. The region is located on the fifth (surface buried) and sixth (surface mostly buried) anti-parallel β-strands; and only the final aa (161–170) are surface exposed. The fourth conserved cluster, ED.4, (16 aa-length) with a percent identity score of 96,96% is comprised by the residues E172-to-N188, which form the long helix α5. The sites 175, 177, 181, 183–184, 186–188 have a percent identity score above 99,9%. This helix is surrounded by the seven antiparallel β-strands and the helix α6. The residues 171–173, 175–176, 180, 182 and 186–187

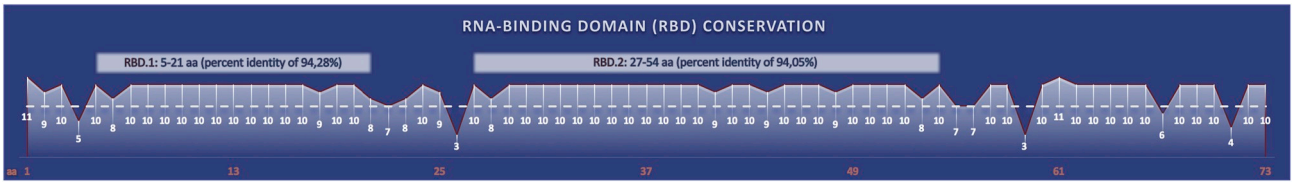


Fig. 2. Representation of the conservation-score distribution of the RNA-Binding Domain (RBD) residues. The numbers in white, below each aa position, represent the conservation score of each individual amino acid, on a scale varying from 0 (most variable) to 11 (most conserved); a score = 11 indicates 100% identity across all human-infecting IAV subtypes. The regions composed of ≥10 continuous conserved amino acids (score > 7) are displayed above the conservation-score bars, where the mean percent identity of the considered region was calculated based on the amino acid frequency information at each particular position.

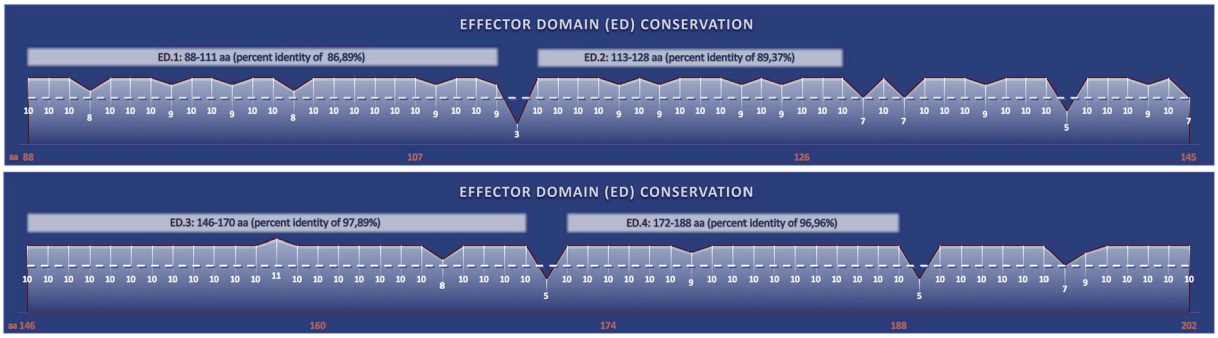


Fig. 3. Representation of the conservation-score distribution of the Effector Domain (ED) residues. The numbers in white, below each aa position, each site represent the conservation score of each individual amino acid, on a scale varying from 0 (most variable) to 11 (most conserved); a score = 11 indicates 100% identity across all human-infecting IAV subtypes. The regions composed of ≥10 continuous conserved amino acids (score > 7) are displayed above the conservation-score bars, where the mean percent identity of the considered region was calculated based on the amino acid frequency information at each particular position.

are surface exposed, while the remaining aa form an interface with the antiparallel β-strands. In the distal part of the ED, two additional areas of 6-residue length each are comprised by highly conserved residues: N190-to-S195 and N197-to-A202, respectively. The first area is located on the strand β7 and the second area belongs to the helix α6 of the NS1 protein.

2.3. Druggability studies

Druggable pockets and sites for the NS1-ED were predicted based on the strategy described in the Materials and Methods. The druggability prediction based on the DoGSiteScorer (DGSS) server is described here in more detail. Only pockets comprised by more than 10 amino acids and with a druggability score ≥0,4 were considered. Druggable pockets tend to be deeper, complex in shape and have a large volume and high enclosure (Perola et al., 2012). The size of the pocket (> 10 residues) is an important descriptor, since small cavities do not favour a proper interaction between the target and the ligand. Pockets which are not able to bind a ligand are described as “decoy pockets” and are usually composed of < 14 residues (Gao and Skolnick, 2013; Hajduk et al., 2005).

Considering the three NS1-ED conformation groups (ED monomers, helix-helix ED dimers and strand-strand ED dimers), twenty-five potential pockets were identified using the DoGSiteScorer server. Some pockets are shared by or in similar regions among the different NS1 conformations. The Supplemental Tables S–2 (supplementary data) displays the total of pockets identified for each NS1-ED conformation group, along with the pocket descriptors and druggability information. The spatial arrangement of these pockets based on the DoGSiteSocrer algorithm is shown in Table 2.

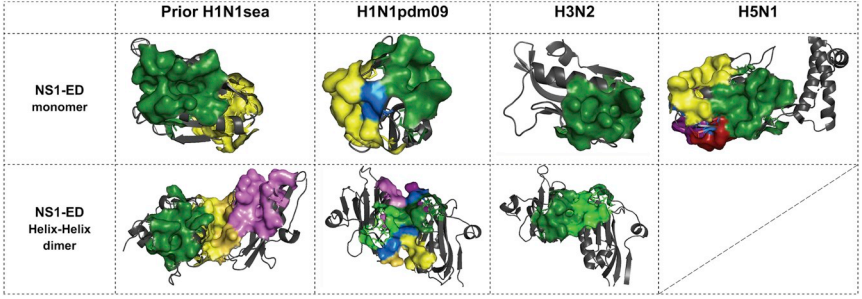
The highest ranked druggable pockets predicted for the NS1-ED monomer were pockets P1 and P2 for the 3O9U structure, with a druggability score of 0,79 and 0,70, respectively. The pocket P1 is composed of 32 amino acids with a conservation score ≥9, except for residues 145 and 171. The pocket comprises the strands β3-β6, the major helix α5 and the minor helix α6. The pocket P2 has 21 conserved amino acids (scores 8–10) and It is comprised by the antiparallel β3-β6 strand extremities.

The highest ranked druggable pockets predicted for the NS1 ED helix-helix dimer interface were pockets P1 for the 3EE9 and P2 for the 3O9U structures, with a druggability score of 0,82 and 0,78, respectively. The pocket P1 (3EE9 structure) is composed of 32 amino acids (16 aa in each NS1 monomer) acids with a conservation score ≥9. The pocket P2 (3O9U structure) is composed of 28 amino acids with a conservation score ≥9, except for the residue 189. Both pockets are mostly located at the α5-α5’ helix interface.

The highest ranked druggable pockets predicted for the NS1-ED strand-strand dimer interface were both pockets P1 for the 3M5R and 3EE8 structures, with a druggability score of 0,87 and 0,74, respectively. The pocket P1 (3M5R structure) is composed of 36 amino acids which are mostly conserved (scores 8–10), except for residues 145 and 171. The pocket P1 (3EE8 structure) is composed of 26 amino acids which are mostly conserved (scores 8–10) with exception of residues 129, 131 and 191 (score = 7); and residue 171 (score = 5). Both pockets mostly lie at the β1-α5′ and β5-β5’ strand interfaces.

A linear schematic representation of the whole NS1-ED druggability is shown in the Fig. S-1 (supplementary data). This analysis was performed by merging the druggability information from the three bioinformatic tools (SF, DGSS and PDS) for each one of the NS1 conformations groups, independently: ED monomer, helix–helix ED dimer and strand-strand ED dimer. The potential druggable regions or sites were coloured in light blue and dark blue. The analysis has revealed

Table 2
Mapping results of the predicted druggable pockets onto the NS1 crystallographic structures of H1N1, H3N2 and H5N1 IAVs. The protein conformations: ED monomer and helix–helix ED dimer were considered. Top-ranking pockets obtained from DGSS are highlighted in shades of green, blue, purple and red, in descending order of rank, respectively. The complete druggability information along with the DGSS algorithm descriptors for each pocket are shown in Supplemental Tables S–2. The figures were produced with PyMOL Molecular Graphics System (Schrodinger, 2015).



differences in the distribution of druggable amino acid regions in the ED, suggesting distinct modes of NS1-ED:host protein interactions. Regarding the ED monomer conformation, the positions 100-to-105 (located between the helix α4 and the strand β5) were found to be druggable, along with the 143–144 region of the strand β5, the 166–169 region (between the strand β5 and strand β6) and part of the helix α5 (aa: 172, 174–175, 177, 180–181).

The residues 110 and 111 (belonging to the strand β2), part of the helix α5 (residues 179, 180, 182, 184, 186–188) and the 189-to-191 region (between the helix α5 and the strand β7) were highlighted as the most druggable regions in the helix–helix ED dimer conformation.

Considering the strand-strand ED dimer conformation, the positions 91-to-94 (located between the strand β1 and the helix α4); the 127-to-129 region of the strand β4; the 137-to-141 region (located between the strand β4 and the strand β5); part of the strand β5 (142–145, 150–151 residues); the 166-to-169 region (located between the strand β6 and the helix α5); and the residues 171 and 172 of the helix α5 were found to be druggable.

The 166–169 region (between the strand β5 and strand β6) and a few residues belonging to the major helix α5 together with part of the strand β5 seem to be global druggable regions transversal across distinct conformations, since they are shared by at least two of the NS1 conformations.

The last stage of the study consisted of a global comparative analysis of the NS1-ED druggability. The druggability consensus shared by the three NS1 conformation groups and predicted by the triple bioinformatic tool strategy (SF, DGSS and PDS) is shown in Fig. 4, together with the degree of residue conservation. The potential conserved druggable regions are marked with an asterisk and the highest-ranked hot spots are marked with a target.

The first conserved druggable cluster was found at positions 100-to-105. This region was predominantly detected in the pockets predicted for the structures of the NS1-ED monomer conformation. The regions 114–115, 126–128, 150–152 were predominantly found in the pockets predicted for the structures of the NS1-ED monomer and NS1 ED strand-strand dimer conformations. The positions 134 and 198–199 consist of a cluster predominantly found in the pockets predicted for the structures of the NS1-ED monomer and NS1 ED helix-helix dimer conformations. The 107–110, 117–123, 141–144, 155–157, 162–163, 166–170, 172–175 (mainly, 172 and 175 positions), 178–182, 184–185, 187–188, 190–194 and 201–202 clusters, along with the 147 site, are represented in the predicted pockets of all the NS1 conformations. The 148 and 159 positions have low or no representation in the predicted pockets.

Some of these regions are placed close together so they can form larger pockets, in respect to the pocket druggability information described in Supplemental Tables S–2. In this regard, three main consensus pockets (of ≥10 aa) were described based on the comparative study of NS1-ED druggability along with the degree of residue conservation (which were named as “consensus druggable pockets”). A consensus druggable pocket (CDP1) frequently found among the three NS1 conformations is comprised by the residues: 114–115, 134, 141–144, 147, 162–163, 166–170, 172–175, 178–182, 194, 198–199, 201–202. The consensus druggable pocket 2 (CDP2) was identified in all the NS1 conformations and comprehends the residues: 108–110, 117–123, 128, 157, 180–181, 184–185, 187–188, 190. A variant of this pocket, the consensus druggable pocket 3 (CDP3), was only found in the NS1 monomer, which includes the residues 100-to-105, as follows: 100–105, 116–123, 126, 151, 155–157, 184–185. A small pocket (SP) comprised by the residues 127–128, 150–152, 191, 193, along with other residues which do not belong to the final druggability consensus, was also found in the NS1-ED strand-strand dimer conformation.

The top-ranked hot spots for drug targeting identified in the global final analysis consisted of the following 24 residues: W102, M104, Q109-K110, P114, D120-A122, N127-I128, L141, L144, A155, V157, P167-H169, K175, V180, G184, T191, R193, F201-A202 (marked with a “target” in Fig. 4). Additionally to the hot spot residues: 122, 127, 141, 144, 167, 191 and 193 that have been previously described in the literature regarding its role in protein function (Hale et al., 2008; Li et al., 2006; Min et al., 2007; Smelkinson et al., 2017; Zhu et al., 2008), we identify, in these study, 9 residues in ligand binding sites previously reported *in silico* by Darapaneni et al. (coloured in orange in Fig. 4) (Darapaneni et al., 2009); and 8 new potential hot spot residues (P114, I128, A155, G168, H169, K175, F201, A202) which, to our knowledge, have not been described before, regarding its druggability and individual role in human IAV (coloured in blue in Fig. 4). The residues 102, 104, 120 and 157 form part of ligand binding site 1; and the residues 109, 110, 120, 121, 180 and 184 form part of ligand binding site

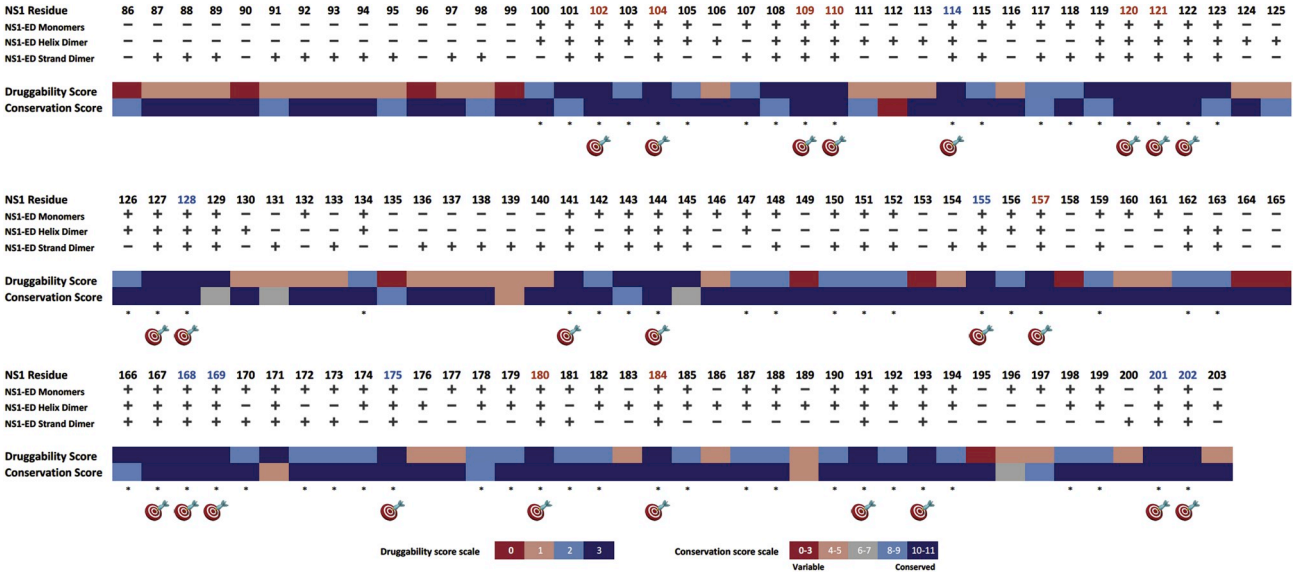


Fig. 4. Overall alignment of NS1-ED druggability along with the conservation scores for each residue position. The druggability prediction was based on the descriptors algorithm of each pocket bioinformatics tool: SF, DGSS and PDS; and for all the predicted NS1 conformations (monomer, helix-helix dimer and strand-strand dimer). The potential conserved druggable regions/sites are marked with an asterisk and the top hot spots are marked with a target. The residue numbers of the 8 new potential hot spot residues, which have not been described in the literature, are coloured in blue. The residues numbers of the 9 residues within the predicted ligand binding sites previously reported by Darapaneni et al. (which overlap with the residues identified in this study) are coloured in orange.

5 reported by Darapaneni et al. (2009). These residues should be further explored *in vitro* and *in vivo* regarding its individual role in protein function or structure on human-infecting IVA. These sites are mostly surface exposed, within pockets with high enclosure, and may represent attractive targets for pharmacological modulation since they potentially establish key interactions with host proteins or other viral factors; or might play a role in NS1 oligomerization, considering its location and the amino acid physical and chemical properties.

We have demonstrated here that highly conservation regions frequently overlap with potential binding sites/residues (characterized with high druggability scores), rendering the NS1 a very good drug-target for small-molecule antiviral drug discovery. For instance, targeting the ED dimerization and oligomerization (mainly mediated by W187), the binding pocket critical for CPSF30 binding (mediated by the residues F103, M106, E186 and W187) the NS1-p85β interaction (mediated by Y89, M93, I95, M98, N133, E142, V145, P162 and P164) or potential binding sites near the RNA-Binding Region (R35, R38, K41, and R46) or involved in RBD dimerization (R35 and R46) have been reported as valuable promising strategies for antiviral design (Das et al., 2008; Engel, 2013; Hale et al., 2010a; Hsu, 2018; Jiao et al., 2008; Kochs et al., 2007; Lalime and Pekosz, 2014; Lopes et al., 2017; Marc, 2014; Schierhorn et al., 2017). The current analysis has revealed additional potential targets within the three main consensus druggable pockets, other than the NS1-RBD:RNA, NS1-ED:CPSF30 and NS1-ED:p85β interaction regions reported in the literature (Engel, 2013; Hale et al., 2010a; Hsu, 2018).

A previous exquisite study based on 2620 NS1 sequences from all influenza subtypes until 2008, carried out by Darapaneni et al., has employed similar approaches in order to identify conserved binding sites predicted for the NS1 H5N1 virus structure, towards the development of anti-NS1 drugs (Darapaneni et al., 2009).

To our knowledge, there has been no published study to date, specifically addressing the conservation and druggability of NS1 protein in such a wide range of NS1 full-length sequences (n = 28392), until 2018, which finally includes the H1N1pdm09 virus and the newly emerging zoonotic H7N9 virus; together with the crystallographic structures of all available human-infecting influenza subtypes (prior H1N1sea, H1N1pdm09, H3N2 and H5N1). Additionally, this is the first approach where distinct bioinformatics tools and algorithms were compared and employed to integrate the conservation information with the druggability prediction for each site within the protein structure, and for all the predicted NS1 conformations (monomer, helix-helix dimer and strand-strand dimer). In this context, all NS1 conformations were taking into account in this work (contrarily to previous studies which only consider the monomeric state of the protein) (Darapaneni et al., 2009; Nayak et al., 2014) because not only the NS1-RBD dimerization is essential for RNA binding, but the effector domain can also dimerize and oligomerize to enable NS1 pleiotropic functions in a spatial and temporal fashion (Aramini et al., 2014). In fact, the homodimeric state of NS1-ED is suggested to be the predominant form at later stages of influenza infection (via either helix-helix or strand-strand dimer conformation) (Aramini et al., 2014; Hale, 2014), so it is essential to be included in NS1 structural studies.

To date, the available full-length NS1 crystal structures are limited to the H5N1 and H6N6 subtypes (Bornholdt and Prasad, 2008; Carrillo et al., 2014). Applying the methodologies used in this study to the dimerized full-length NS1 protein of H1N1pdm09, H3N2 and other IAV subtypes, along with NS1:host-proteins or NS1-RNA complexes, would significantly help to clarify the comprehension of NS1 pleiotropic and multistructural features.

In silico computational methodologies accelerate the drug design process and guide the experimental efforts, but they have limitations considering the lack of experimental validation (Cichonska et al., 2017; Sliwoski et al., 2014). Hence, the hypotheses presumed in this study should be strengthened by experimental data.

3. Conclusion

In this study, we report a promising antiviral strategy directed to highly conserved druggable key NS1 regions as a result of an exhaustive large-scale sequence analysis and structure characterization of NS1 protein across multiple human-infecting influenza A subtypes, which promotes target identification efficiency. To the best of our knowledge, this is the most comprehensive study regarding the conservation and druggability of influenza A NS1 performed to date.

The data surveyed here is a starting point to further explore the molecular mechanisms underpinning NS1 multifunction nature, with the aim of identifying the most promising cellular and viral pathways and the corresponding NS1:host-protein interaction complexes that can, in turn, be investigated and targeted for novel anti-NS1 strategies. In a broader-sense, this research also contributes to a deeper comprehension into the evolutionary dynamics of influenza NS1.

We have identified 8 new potential conserved hot spot residues in the NS1 protein which have not been described before in the literature, additionally to other 9 residues in ligand binding sites previously reported by Darapanei et al.; and also *7* residues which have been described in the literature as playing an important role in protein structure or function. These surface-distributed sites represent attractive targets for pharmacological modulation, rendering the NS1 a promising target candidate for molecular medicine. We preconize that an anti-influenza strategy targeting potential functionally or structurally highly conserved sites, in a protein exclusively expressed in the infected cell, is less prone to the emergence of resistant variants and are expected to be useful against all influenza A viruses affecting the human host.

Our study provides a robust roadmap and a structure-based rationale that have the potential to accelerate NS1 target validation. The findings from this project contribute to enlighten the clinical applicability of the influenza A virus NS1 as an emerging therapeutic strategy. In this context, this study lays the basis for future work in the discovery and design of antiviral drugs targeting the NS1 protein.

4. Forthcoming research

Further studies are important to explore and characterize the structure–function relationship of the disclosed highly conserved residues and identify the role for individual sites.

Considering that NS1 is a multifunctional protein, which interacts with multiple host proteins and viral factors, exploring individual sites regarding the structure–function relationship studies, can make the interpretation of these data very complex, since a single (loss-of-function) mutation might exert more than one consequence regarding NS1 function during IAV infection. In this context, it would be interesting to explore the corresponding cellular and viral pathways and the underpinning mechanisms that are affecting the effective viral replication for a specific site. Hereof, the results could also contribute to a comprehensively understanding of the infection pathogenesis and dynamics of virus replication regarding NS1 multifunction.

Further studies regarding NS1 druggability should also be extended to the RBD domain.

The strategies used for NS1 protein can be broadened applied to other influenza proteins for a structure-based prediction of druggable targets in the direction of new antiviral therapy approaches.

5. Material and methods

5.1. Dataset construction and sequence analysis

NS nucleotide sequences of worldwide circulating viruses – available since 1918 until February 2018 – from human-infecting influenza A subtypes were obtained from the NCBI Influenza Virus database at www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html, GISAID's EpiFlu™ database at www.gisaid.org and Influenza Research Database at https://

www.fludb.org/brc/home.spg?decorator=influenza (Bao et al., 2008; Shu and McCauley, 2017; Zhang et al., 2017).

All available sequences with a complete coding region were collected for the study. Duplicated sequences between the sources and sequences containing degenerate nucleotides were removed. Additionally, 51 unpublished sequences of viruses circulating in Portugal during 2009–2010 and 2010–2011 seasons were included to the final dataset. The nucleotide sequence alignment was performed using the Clustal Omega webserver available at https://www.ebi.ac.uk/Tools/msa/clustalo/(Kumar et al., 2016; Sievers et al., 2011).

The study of protein length variation (LV) distributions among each influenza virus subtype was performed using MEGA7 (Kumar et al., 2016).

5.2. Conservation analysis

The amino acid conservation in the multiple sequence alignment (MSA) was calculated using the Valdar scoring method from the Jalview AACons Web server (version 2.10.4) (Valdar, 2002; Waterhouse et al., 2009). The weighted scores incorporated into the Valdar conservation scoring method consider the sequence redundancy in the multiple sequence alignment, since the sequence weighting attributes higher weights to sequences that present greater diversity from the rest of the alignment. In this way, it normalizes against redundancy and bias in the MSA (reducing the effect of bias sampling and penalizing gaps) with no loss of evolutionary information (Valdar, 2002). This scoring system has been previously applied with success by other authors regarding the conservation analysis of influenza virus proteins other than NS1, since this system has been suggested to be essential for a proper conservation score analysis of influenza virus sequences (Kukol, 2017; Patel and Kukol, 2017). The conservation degree was given as a numerical score, with a range between 0 (high variable site) and 11 (highly conserved site), for each amino acid site in the protein (Valdar, 2002; Waterhouse et al., 2009). Residues with conservation score of 11 corresponded to absolute conserved sites (100% amino acid identity); residues with conservation score ≥ 10 were considered highly conserved; residues with conservation score of 7–9 were considered conserved; and residues with conservation score ≤ 4 were considered variable (Livingstone and Barton, 1993). Considering that the LR and the CTT contain several gaps and deletions, both regions were considered *a priori* as variable regions. The distribution of percent identity of all conserved regions (containing ≥ 10 aa) was also studied. The amino acid frequency percentage for each position and per region within the NS1 protein were calculated using the computer program DAMBE version 6.4.40 (Xia and Xie, 2001).

5.3. Protein structure

Crystallographic structures of influenza NS1 protein were obtained from the RCSB Protein Data Bank (www.rcsb.org) (Berman et al., 2000). Reasonable X-ray crystallographic structures of the prior seasonal H1N1 influenza (PDB entries 2GX9, 3O9U) (Kerry et al., 2011); H1N1pdm09 influenza (PDB entry 3M5R); H3N2 influenza (PDB entry 3EE9) (Xia et al., 2009); and the H5N1 influenza (PDB entries 3F5T – full length; and 3P38) (Bornholdt and Prasad, 2008) structures were selected for the study. The crystallographic three-dimensional structures were accurately prepared (including removing water and bounded ligands, protonation and energy minimization) using MOE 2015.10001 program (Molecular Operating Environment (MOE), 2016).

5.4. Druggability

The druggable sites in the effector domain of NS1 protein were predicted using the webservers DoGSiteScorer at https://proteins.plus and PockDrug-Server at http://pockdrug.rpbs.univ-paris-diderot.fr/; and the commercial software MOE-SiteFinder from Chemical Computing Group (Hussein et al., 2015; Molecular Operating Environment (MOE), 2016; Volkamer et al., 2012).

The commercial tool **SF** of MOE 2015.1**0001** software is based on a geometry-based method. The top ranked sites with a positive score, according to the Propensity for Ligand Binding – PLB index and the hydrophobic contact count, were considered druggable (Molecular Operating Environment (MOE), 2016).

DGSS server (https://proteins.plus) is based on a grid-based method incorporated in a support vector machine model for druggability predictions. The algorithm is based on geometric (size, shape) and physicochemical descriptors and provides a druggability score between the values: 0 (undruggable) to 1 (druggable), for each pocket. Pockets with a druggability score ≥ 0.4 were considered druggable (Volkamer et al., 2012).

PDS (http://pockdrug.rpbs.univ-paris-diderot.fr) is based on a geometry-based method (Hussein et al., 2015). The PDS tool retains the best combinations of three pocket properties: geometry, hydrophobicity, and aromaticity. Pockets with a druggability score ≥ 0.5 were considered druggable (Hussein et al., 2015).

Initially, the druggability prediction for the NS1 conformation groups: (1) ED monomers (PDB IDs: 3EE9, 3F5T, 3M5R and 3O9U); (2) the helix–helix ED dimers (PDB IDs: 3EE9, 3M5R and 3O9U); and (3) the strand-strand ED dimers (PDB IDs: 2GX9, 3M5R and 3P38) has been individually studied using one by one of the three bioinformatic tools: (a) SF, (b) DGSS and (c) PDS. Subsequently, a comparative analysis was performed by merging the druggability information from the three bioinformatic tools for each one of the NS1 conformation groups.

The last stage of the study consisted of a global comparative analysis, including the most prevailing druggable sites from the three bioinformatic tools together, and shared by all the three NS1 conformation groups, in order to produce a final druggability prediction - druggability consensus. The druggability consensus was manually analysed in parallel with the conservation information in order to identify druggable binding sites/pockets or potential hot spots residues that overlap or are spatially close to the conserved NS1 regions/residues (which were named as “consensus druggable pockets”).

The coordinates of the highly score druggable pockets were mapped onto the three-dimensional NS1 structures. The figures were prepared with PyMOL Molecular Graphics System, version 1.7.6 (Schrödinger, LLC; https://pymol.org). (Schrodinger, 2015).

Funding

This work was supported by the grant PD/BD/128402/2017 from Fundação para a Ciência e a Tecnologia (FCT) PhD Programme in Medicines and Pharmaceutical Innovation (i3DU) and by FCT grants: PTDC/QUI-OUT/32572/2017 and PTDC/SAU-INF/30729/2017.

Declarations of interest

None.

Acknowledgments

The authors would like to thank to all authors, originating and submitting laboratories that share their sequences through the GISAID EpiFlu™, Influenza Research Database and NCBI Influenza Virus database databases.

The authors thank to Maria de São José Alexandre (Universidade do Porto) and Rui M.M. Brito (Centro de Química de Coimbra, Universidade de Coimbra) for the guidance given as supervisors of the Ph.D. work of J.Trigueiro-Louro.

The authors also thank Rita Guedes Lab Group (Medicinal Chemistry Unit, iMed.ULisboa, FFUL), namely Bruno Victor, Filipa Gomes, Rita Acúrcio and Sara Garcia not only for providing the required computational equipment but also for sharing their expertise

regarding some computational methodologies applied to this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virol.2019.04.009.

References

Abdelwhab, E.M., Veits, J., Mettenleiter, T.C., 2013. Avian influenza virus NS1: a small protein with diverse and versatile functions. Virulence 4, 583–588. https://doi.org/10.4161/viru.26360.

Aramini, J.M., Hamilton, K., Ma, L.C., Swapna, G.V.T., Leonard, P.G., Ladbury, J.E., Krug, R.M., Montelione, G.T., 2014. (19)F NMR reveals multiple conformations at the dimer interface of the nonstructural protein 1 effector domain from influenza A virus. Structure 22, 515–525. https://doi.org/10.1016/j.str.2014.01.010.

Bao, Y., Bolotov, P., Dernovoy, D., Kiryutin, B., Zaslavsky, L., Tatusova, T., Ostell, J., Lipman, D., 2008. The influenza virus resource at the national center for biotechnology information. J. Virol. 82, 596–601. https://doi.org/10.1128/JVI.02005-07.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein Data Bank. Nucleic Acids Res. 28, 235–242.

Bornholdt, Z.A., Prasad, B.V., 2008. X-ray structure of NS1 from a highly pathogenic H5N1 influenza virus. Nature 456, 985–988. https://doi.org/10.1038/nature07444.

Carrillo, B., Choi, J.M., Bornholdt, Z.A., Sankaran, B., Rice, A.P., Prasad, B.V., 2014. The influenza A virus protein NS1 displays structural polymorphism. J. Virol. 88, 4113–4122. https://doi.org/10.1128/JVI.03692-13.

Cichonska, A., Ravikumar, B., Parri, E., Timonen, S., Pahikkala, T., Airola, A., Wennerberg, K., Rousu, J., Aittokallio, T., 2017. Computational-experimental approach to drug-target interaction mapping: a case study on kinase inhibitors. PLoS Comput. Biol. 13 e1005678–e1005678. https://doi.org/10.1371/journal.pcbi.1005678.

Cox, N.J., 1999. Prevention and control of influenza. Lancet 354 (Suppl. l) SIV30.

Darapaneni, V., Prabhaker, V.K., Kukol, A., 2009. Large-scale analysis of influenza A virus sequences reveals potential drug target sites of non-structural proteins. J. Gen. Virol. 90, 2124–2133. https://doi.org/10.1099/vir.0.011270-0.

Das, K., Ma, L.C., Xiao, R., Radvansky, B., Aramini, J., Zhao, L., Marklund, J., Kuo, R.L., Twu, K.Y., Arnold, E., Krug, R.M., Montelione, G.T., 2008. Structural basis for suppression of a host antiviral response by influenza A virus. Proc. Natl. Acad. Sci. U. S. A. 105, 13093–13098. https://doi.org/10.1073/pnas.0805213105.

Dundon, W.G., Capua, I., 2009. A closer look at the NS1 of influenza virus. Viruses Basel 1, 1057–1072. https://doi.org/10.3390/v1031057.

Engel, D.A., 2013. The influenza virus NS1 protein as a therapeutic target. Antivir. Res. 99, 409–416. https://doi.org/10.1016/j.antiviral.2013.06.005.

Gao, M., Skolnick, J., 2013. APoc: large-scale identification of similar protein pockets. Bioinformatics 29, 597–604. https://doi.org/10.1093/bioinformatics/btt024.

Garcia-Sastre, A., 2017. Ten strategies of interferon evasion by viruses. Cell Host Microbe 22, 176–184. https://doi.org/10.1016/j.chom.2017.07.012.

Ghebrehewet, S., MacPherson, P., Ho, A., 2016. Influenza. BMJ 355, i6258. https://doi.org/10.1136/bmj.i6258.

Golebiewski, L., Liu, H., Javier, R.T., Rice, A.P., 2011. The avian influenza virus NS1 ESEV PDZ binding motif associates with Dlg1 and Scribble to disrupt cellular tight junctions. J. Virol. 85, 10639–10648. https://doi.org/10.1128/JVI.05070-11.

Hajduk, P.J., Huth, J.R., Tse, C., 2005. Predicting protein druggability. Drug Discov. Today 10, 1675–1682. https://doi.org/10.1016/S1359-6446(05)03624-X.

Hale, B.G., 2014. Conformational plasticity of the influenza A virus NS1 protein. J. Gen. Virol. 95, 2099–2105. https://doi.org/10.1099/vir.0.066282-0.

Hale, B.G., Kerry, P.S., Jackson, D., Precious, B.L., Gray, A., Killip, M.J., Randall, R.E., Russell, R.J., 2010a. Structural insights into phosphoinositide 3-kinase activation by the influenza A virus NS1 protein. Proc. Natl. Acad. Sci. U. S. A. 107, 1954–1959. https://doi.org/10.1073/pnas.0910715107.

Hale, B.G., Knebel, A., Botting, C.H., Galloway, C.S., Precious, B.L., Jackson, D., Elliott, R.M., Randall, R.E., 2009. CDK/ERK-mediated phosphorylation of the human influenza A virus NS1 protein at threonine-215. Virology 383, 6–11. https://doi.org/10.1016/j.virol.2008.10.002.

Hale, B.G., Randall, R.E., Ortin, J., Jackson, D., 2008. The multifunctional NS1 protein of influenza A viruses. J. Gen. Virol. 89, 2359–2376. https://doi.org/10.1099/vir.0.2008/004606-0.

Hale, B.G., Steel, J., Manicassamy, B., Medina, R.A., Ye, J.Q., Hickman, D., Lowen, A.C., Perez, D.R., Garcia-Sastre, A., 2010b. Mutations in the NS1 C-terminal tail do not enhance replication or virulence of the 2009 pandemic H1N1 influenza A virus. J. Gen. Virol. 91, 1737–1742. https://doi.org/10.1099/vir.0.020925-0.

Hrincius, E.R., Dierkes, R., Anhlan, D., Wixler, V., Ludwig, S., Ehrhardt, C., 2011. Phosphatidylinositol-3-kinase (PI3K) is activated by influenza virus vRNA via the pathogen pattern receptor Rig-I to promote efficient type I interferon production. Cell Microbiol. 13, 1907–1919. https://doi.org/10.1111/j.1462-5822.2011.01680.x.

Hsiang, T.Y., Zhou, L., Krug, R.M., 2012. Roles of the phosphorylation of specific serines and threonines in the NS1 protein of human influenza A viruses. J. Virol. 86, 10370–10376. https://doi.org/10.1128/JVI.00732-12.

Hsu, A.C.Y., 2018. Influenza virus: a master tactician in innate immune evasion and novel therapeutic interventions. Front. Immunol. 9. https://doi.org/10.3389/fimmu.2018.00743.

Hurt, A.C., Besselaar, T.G., Daniels, R.S., Ermetal, B., Fry, A., Gubareva, L., Huang, W., Lackenby, A., Lee, R.T., Lo, J., Maurer-Stroh, S., Nguyen, H.T., Pereyaslov, D., Rebelo-de-Andrade, H., Siqueira, M.M., Takashita, E., Tashiro, M., Tilmanis, D., Wang, D., Zhang, W., Meijer, A., 2016. Global update on the susceptibility of human influenza viruses to neuraminidase inhibitors, 2014-2015. Antivir. Res. 132, 178–185. https://doi.org/10.1016/j.antiviral.2016.06.001.

Hussein, H.A., Borrel, A., Geneix, C., Petitjean, M., Regad, L., Camproux, A.C., 2015. PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. Nucleic Acids Res. 43, W436–W442. https://doi.org/10.1093/nar/gkv462.

Jackson, D., Hossain, M.J., Hickman, D., Perez, D.R., Lamb, R.A., 2008. A new influenza virus virulence determinant: the NS1 protein four C-terminal residues modulate pathogenicity. Proc. Natl. Acad. Sci. U. S. A. 105, 4381–4386. https://doi.org/10.1073/pnas.0800482105.

Jiao, P.R., Tian, G.B., Li, Y.B., Deng, G.H., Jiang, Y.P., Liu, C., Liu, W.L., Bu, Z.G., Kawaoka, Y., Chen, H.L., 2008. A single-amino-acid substitution in the NS1 protein changes the pathogenicity of H5N1 avian influenza viruses in mice (vol 82, pg 1146, 2008). J. Virol. 82, 4190. https://doi.org/10.1128/Jvi.00376-08.

Kerry, P.S., Ayllon, J., Taylor, M.A., Hass, C., Lewis, A., Garcia-Sastre, A., Randall, R.E., Hale, B.G., Russell, R.J., 2011. A transient homotypic interaction model for the influenza A virus NS1 protein effector domain. PLoS One 6 e17946. https://doi.org/10.1371/journal.pone.0017946.

Kirchmair, J., Distinto, S., Liedl, K.R., Markt, P., Rollinger, J.M., Schuster, D., Spitzer, G.M., Wolber, G., 2011. Development of anti-viral agents using molecular modeling and virtual screening techniques. Infect. Disord. - Drug Targets 11, 64–93.

Kleinpeter, A.B., Jureka, A.S., Falahat, S.M., Green, T.J., Petit, C.M., 2018. Structural analyses reveal the mechanism of inhibition of influenza virus NS1 by two antiviral compounds. J. Biol. Chem. 293, 14659–14668. https://doi.org/10.1074/jbc.RA118.004012.

Klepser, M.E., 2014. Socioeconomic impact of seasonal (epidemic) influenza and the role of over-the-counter Medicines. Drugs 74, 1467–1479. https://doi.org/10.1007/s40265-014-0245-1.

Kochs, G., Garcia-Sastre, A., Martinez-Sobrido, L., 2007. Multiple anti-interferon actions of the influenza A virus NS1 protein. J. Virol. 81, 7011–7021. https://doi.org/10.1128/Jvi.02581-07.

Krammer, F., Smith, G.J.D., Fouchier, R.A.M., Peiris, M., Kedzierska, K., Doherty, P.C., Palese, P., Shaw, M.L., Treanor, J., Webster, R.G., Garcia-Sastre, A., 2018. Influenza. Nat. Rev. Dis. Prim. 4, 3. https://doi.org/10.1038/s41572-018-0002-y.

Krug, R.M., 2015. Functions of the influenza A virus NS1 protein in antiviral defense. Curr. Opin. Virol. 12, 1–6. https://doi.org/10.1016/j.coviro.2015.01.007.

Kukol, A., 2017. Combining Biology and Chemistry to Target the Flu Virus. Lap Lambert Academic Publishing, Beau Bassin.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874. https://doi.org/10.1093/molbev/msw054.

Kuo, R.L., Li, L.H., Lin, S.J., Li, Z.H., Chen, G.W., Chang, C.K., Wang, Y.R., Tam, E.H., Gong, Y.N., Krug, R.M., Shih, S.R., 2016. Role of N Terminus-Truncated NS1 proteins of influenza a virus in inhibiting IRF3 activation. J. Virol. 90, 4696–4705. https://doi.org/10.1128/Jvi.02843-15.

Lalime, E.N., Pekosz, A., 2014. The R35 residue of the influenza A virus NS1 protein has minimal effects on nuclear localization but alters virus replication through disrupting protein dimerization. Virology 458, 33–42. https://doi.org/10.1016/j.virol.2014.04.012.

Li, S.D., Min, J.Y., Krug, R.M., Sen, G.C., 2006. Binding of the influenza A virus NS1 protein to PKR mediates the inhibition of its activation by either PACT or double-stranded RNA. Virology 349, 13–21. https://doi.org/10.1016/j.virol.2006.01.005.

Livingstone, C.D., Barton, G.J., 1993. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. Comput. Appl. Biosci. 9, 745–756.

Lopes, A.M., Domingues, P., Zell, R., Hale, B.G., 2017. Structure-guided functional annotation of the influenza a virus NS1 protein reveals dynamic evolution of the p85beta-binding site during circulation in humans. J. Virol. 91. https://doi.org/10.1128/JVI.01081-17.

Lv, W., Xu, Y., Guo, Y., Yu, Z., Feng, G., Liu, P., Luan, M., Zhu, H., Liu, G., Zhang, M., Lv, H., Duan, L., Shang, Z., Li, J., Jiang, Y., Zhang, R., 2015. The Drug Target Genes Show Higher Evolutionary Conservation than Non-target Genes. Oncotarget. https://doi.org/10.18632/oncotarget.6755.

Mallipeddi, P.L., Kumar, G., White, S.W., Webb, T.R., 2014. Recent advances in computer-aided drug design as applied to anti-influenza drug discovery. Curr. Top. Med. Chem. 14, 1875–1889.

Marc, D., 2016. Stop-codon variations in non-structural protein NS1 of avian influenza viruses. Virulence 7, 498–501. https://doi.org/10.1080/21505594.2016.1175802.

Marc, D., 2014. Influenza virus non-structural protein NS1: interferon antagonism and beyond. J. Gen. Virol. 95, 2594–2611. https://doi.org/10.1099/vir.0.069542-0.

Min, J.Y., Li, S.D., Sen, G.C., Krug, R.M., 2007. A site on the influenza A virus NS1 protein mediates both inhibition of PKR activation and temporal regulation of viral RNA synthesis. Virology 363, 236–243. https://doi.org/10.1016/j.virol.2007.01.038.

Mohs, R.C., Greig, N.H., 2017. Drug discovery and development: role of basic biological research. Alzheimers Dement. (N Y) 3, 651–657. https://doi.org/10.1016/j.trci.2017.10.005.

Molecular Operating Environment (MOE), 2013. 2013.08, 2016. Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7. Mol. Oper. Environ. (MOE), 2013.08; Chem. Comput. Gr. Inc., 1010 Sherbooke St. West, Suite #910, Montr. QC, Canada, H3A 2R7.

Nayak, M.K., Agrawal, A.S., Bose, S., Naskar, S., Bhowmick, R., Chakrabarti, S., Sarkar, S., Chawla-Sarkar, M., 2014. Antiviral activity of baicalin against influenza virus H1N1-pdm09 is due to modulation of NS1-mediated cellular innate immune responses. J.

Antimicrob. Chemother. 69, 1298–1310. https://doi.org/10.1093/jac/dkt534.
Obenauer, J.C., Denson, J., Mehta, P.K., Su, X., Mukatira, S., Finkelstein, D.B., Xu, X., Wang, J., Ma, J., Fan, Y., Rakestraw, K.M., Webster, R.G., Hoffmann, E., Krauss, S., Zheng, J., Zhang, Z., Naeve, C.W., 2006. Large-scale sequence analysis of avian influenza isolates. Science 311 (80), 1576–1580. https://doi.org/10.1126/science.1121586.
Patel, H., Kukol, A., 2017. Evolutionary conservation of influenza A PB2 sequences reveals potential target sites for small molecule inhibitors. Virology 509, 112–120. https://doi.org/10.1016/j.virol.2017.06.009.
Penttinen, P., Catchpole, M., 2016. ECDC expert opinion on efficacy and effectiveness of neuraminidase inhibitors published for public consultation. Influenza and Other Respir. Viruses 10, 152–153. https://doi.org/10.1111/irv.12377.
Perola, E., Herman, L., Weiss, J., 2012. Development of a rule-based method for the assessment of protein druggability. J. Chem. Inf. Model. 52, 1027–1038. https://doi.org/10.1021/ci200613b.
Putri, W.C.W.S., Muscatello, D.J., Stockwell, M.S., Newall, A.T., 2018. Economic burden of seasonal influenza in the United States. Vaccine 36, 3960–3966. https://doi.org/10.1016/j.vaccine.2018.05.057.
Robb, N.C., Jackson, D., Vreede, F.T., Fodor, E., 2010. Splicing of influenza A virus NS1 mRNA is independent of the viral NS1 protein. J. Gen. Virol. 91, 2331–2340. https://doi.org/10.1099/vir.0.022004-0.
Schierhorn, K.L., Jolmes, F., Bespalowa, J., Saenger, S., Peteranderl, C., Dzieciolowski, J., Mielke, M., Budt, M., Pleschka, S., Herrmann, A., Herold, S., Wolff, T., 2017. Influenza a virus virulence depends on two amino acids in the N-terminal domain of its NS1 protein to facilitate inhibition of the RNA-dependent protein kinase PKR. J. Virol. 91. https://doi.org/10.1128/JVI.00198-17.
Schrodinger, L.L.C., 2015. The PyMOL Molecular Graphics System. Version 1.8.
Schuhmacher, A., Gassmann, O., Hinder, M., 2016. Changing R&D models in research-based pharmaceutical companies. J. Transl. Med. 14, 105. https://doi.org/10.1186/s12967-016-0838-4.
Shu, Y., McCauley, J., 2017. GISAID: global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 22. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494.
Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539. https://doi.org/10.1038/msb.2011.75.
Sliwoski, G., Kothiwale, S., Meiler, J., Lowe Jr., E.W., 2014. Computational methods in drug discovery. Pharmacol. Rev. 66, 334–395. https://doi.org/10.1124/pr.112.007336.
Smelkinson, M.G., Guichard, A., Teijaro, J.R., Malur, M., Loureiro, M.E., Jain, P., Ganesan, S., Zuniga, E.I., Krug, R.M., Oldstone, M.B., Bier, E., 2017. Influenza NS1 directly modulates Hedgehog signaling during infection. PLoS Pathog. 13. https://doi.org/10.1371/journal.ppat.1006588.
Takashita, E., Meijer, A., Lackenby, A., Gubareva, L., Rebelo-de-Andrade, H., Besselaar, T., Fry, A., Gregory, V., Leang, S.K., Huang, W., Lo, J., Pereyaslov, D., Siqueira, M.M., Wang, D., Mak, G.C., Zhang, W., Daniels, R.S., Hurt, A.C., Tashiro, M., 2015. Global update on the susceptibility of human influenza viruses to neuraminidase inhibitors, 2013-2014. Antivir. Res. 117, 27–38. https://doi.org/10.1016/j.antiviral.2015.02.003.
Valdar, W.S., 2002. Scoring residue conservation. Proteins 48, 227–241. https://doi.org/10.1002/prot.10146.
Volkamer, A., Kuhn, D., Rippmann, F., Rarey, M., 2012. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. Bioinformatics 28, 2074–2075. https://doi.org/10.1093/bioinformatics/bts310.
Wang, W., Riedel, K., Lynch, P., Chien, C.Y., Montelione, G.T., Krug, R.M., 1999. RNA binding by the novel helical domain of the influenza virus NS1 protein requires its dimer structure and a small number of specific basic amino acids. RNA 5, 195–205.
Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., Barton, G.J., 2009. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191. https://doi.org/10.1093/bioinformatics/btp033.
Xia, S., Robertus, J.D., 2010. X-ray structures of NS1 effector domain mutants. Arch. Biochem. Biophys. 494, 198–204. https://doi.org/10.1016/j.abb.2009.12.008.
Xia, S.L., Monzingo, A.F., Robertus, J.D., 2009. X-ray structure of NS1A effector domain from the influenza A/Udorn/72 virus. FASEB J. 23.
Xia, X., Xie, Z., 2001. DAMBE: software package for data analysis in molecular biology and evolution. J. Hered. 92, 371–373. https://doi.org/10.1093/jhered/92.4.371.
Zhang, Y., Aevermann, B.D., Anderson, T.K., Burke, D.F., Dauphin, G., Gu, Z., He, S., Kumar, S., Larsen, C.N., Lee, A.J., Li, X., Macken, C., Mahaffey, C., Pickett, B.E., Reardon, B., Smith, T., Stewart, L., Suloway, C., Sun, G., Tong, L., Vincent, A.L., Walters, B., Zaremba, S., Zhao, H., Zhou, L., Zmasek, C., Klem, E.B., Scheuermann, R.H., 2017. Influenza Research Database: an integrated bioinformatics resource for influenza virus research. Nucleic Acids Res. 45, D466–D474. https://doi.org/10.1093/nar/gkw857.
Zhu, Q.Y., Yang, H.H., Chen, W.Y., Cao, W.Y., Zhonf, G.X., Jiao, P.R., Deng, G.H., Yu, K.Z., Yang, C.L., Bu, Z.G., Kawaoka, Y., Chen, H.L., 2008. A naturally occurring deletion in its NS gene contributes to the attenuation of an H5N1 swine influenza virus in chickens. J. Virol. 82, 220–228. https://doi.org/10.1128/Jvi.00978-07.