Virology 509 (2017) 112–120

Contents lists available at ScienceDirect

Virology

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

Evolutionary conservation of influenza A PB2 sequences reveals potential target sites for small molecule inhibitors

Hershna Patel, Andreas Kukol*

School of Life and Medical Sciences, University of Hertfordshire, Hatfield AL10 9AB, United Kingdom

ARTICLE INFO

Keywords:
Influenza A
PB2
Sequence evolution
Binding site
Virtual screening
Drug discovery
Conservation
Paliperidone

ABSTRACT

The influenza A basic polymerase protein 2 (PB2) functions as part of a heterotrimer to replicate the viral RNA genome. To investigate novel PB2 antiviral target sites, this work identified evolutionary conserved regions across the PB2 protein sequence amongst all sub-types and hosts, as well as ligand binding hot spots which overlap with highly conserved areas. Fifteen binding sites were predicted in different PB2 domains; some of which reside in areas of unknown function. Virtual screening of ~50,000 drug-like compounds showed binding affinities of up to −10.3 kcal/mol. The highest affinity molecules were found to interact with conserved residues including Gln138, Gly222, Ile529, Asn540 and Thr530. A library containing 1738 FDA approved drugs was screened additionally and revealed Paliperidone as a top hit with a binding affinity of −10 kcal/mol. Predicted ligands are ideal leads for new antivirals as they were targeted to evolutionary conserved binding sites.

1. Introduction

The influenza A virus causes one of the most prevalent and significant respiratory viral infections worldwide with previous pandemics resulting in remarkably high fatality rates (Taubenberger and Kash, 2010). This is largely due to continuous genome evolution and the zoonotic nature of the virus which enables rapid transmission of new re-assortant strains (Reperant et al., 2012). The main form of prevention against influenza is annual vaccination; however this may not always guarantee extensive protection or control of the virus (Chambers et al., 2015). Therefore treatment with antiviral drugs such as the neuraminidase inhibitors is heavily relied on, while there is widespread resistance against matrix protein2 (M2) inhibitors. Since the adoption of these drugs, influenza A subtypes in circulation have shown varying levels of sensitivity due to amino acid mutations in the drug target site (Hayden and De Jong, 2011; Samson et al., 2013). For this reason the M2 inhibitors are no longer recommended for clinical use (Harper et al., 2009). Consequently, the discovery and search for novel antivirals which are unlikely to be affected by resistance mutations is a priority, with several candidate inhibitors having emerged from recent studies (reviewed in Naesens et al. (2016); Patel and Kukol (2016)).

The infectious life cycle of the virus requires several functional proteins encoded by eight RNA segments, which are released into the host cell, allowing the virus to replicate its genome and suppress the immune response (Bouvier and Palese, 2008). The influenza A polymerase basic protein 2 (PB2) is encoded by RNA segment one. It is one of the largest influenza proteins consisting of 759 amino acids and is a constituent subunit of the trimeric viral polymerase complex in addition to polymerase basic protein 1 (PB1) and the acidic polymerase (PA). Transcription and replication of the viral genome occurs in the host cell nucleus, and involves a series of stages before translation of viral mRNA in the cytoplasm (Fodor, 2013). During transcription, the PB2 protein is mainly responsible for generating the cap structure for viral mRNA from the 5' end of 7-methyl guanosine triphosphate (mGTP) capped host mRNA. The PB2 'cap snatching' mechanism involves residues between positions 318–482, which recognise methylated guanosine in order to bind the host cell RNA strand. The endonuclease subunit of the PA then cleaves the RNA leaving a 10–13 nucleotide primer to initiate transcription by PB1 (Fodor, 2013). In complex, the N-terminal 249 residues of the PB2 subunit are associated with the C-terminal subunit of PB1, which is a critical interaction to trigger PA endonuclease activity (Sugiyama et al., 2009). A structural study of the PB2 protein from a H5N1 avian virus had found that following translation, the C-terminal domain (residues 536–759) undergoes large conformational re-organisation between open and closed states. This flexibility enables the nuclear localization signal (NLS) peptide in the 686–759 region to bind with host importin-α,

Abbreviations: PB2, polymerase basic protein 2; PB1, polymerase basic protein 1; PA, polymerase acidic; PAINS, pan assay interference compounds; NCI, National Cancer Institute; NCBI, National Centre for Biotechnology Information; NP, nucleoprotein; mGTP, methylguanosine triphosphate; NLS, nuclear localization signalling

* Corresponding author.
E-mail addresses: h.patel28@herts.ac.uk (H. Patel), a.kukol@herts.ac.uk (A. Kukol).

http://dx.doi.org/10.1016/j.virol.2017.06.009
Received 26 April 2017; Received in revised form 7 June 2017; Accepted 9 June 2017

enabling PB2 entry into the target cell nucleus to catalyze further RNA transcription (Das et al., 2010; Delaforge et al., 2015). Other PB2 conformational changes occurring in connection with the cap-snatching mechanism and the kind of RNA bound have also been described in the context of the full polymerase complex (Reich et al., 2014; Thierry et al., 2016).

In addition to mutations in the haemagglutinin (HA) and neuraminidase (NA) proteins, changes in the sequences of polymerase proteins are also considered as major determinants of host range and adaptation (Mehle and Doudna, 2009; Neumann and Kawaoka, 2015). The characteristic PB2 host determining residue at position 627 (with lysine being prevalent in human strains, glutamic acid present in avian strains and serine in bat strains) is situated in a loop region, which along with the cap-binding domain does not make extensive contact with the PB1 and PA subunits (Kuzuhara et al., 2009; Pflug et al., 2014). The 535–684 domain has also been shown to have RNA binding activity which is affected by the E627K mutation and is unrelated to the cap-snatching function (Kuzuhara et al., 2009).

As the PB2 protein plays multiple essential roles in the virus life cycle, it is a valid target for antiviral drugs. Several crystal structures are available in the protein data bank (PDB) for specific PB2 subunit domains in holo and apo forms which can aid with structure based drug discovery studies. Despite some of the PB2 surface area being inaccessible due to trimer assembly, inhibitors of the 'cap snatching' function to prevent capped host mRNA binding have been identified showing potent effects against several influenza strains in vitro (Boyd et al., 2015; Clark et al., 2014; Pautus et al., 2013). The aim of this work was to identify highly conserved regions of PB2 using an algorithm superior to counting conservation from multiple sequence alignments and to identify overlap with potential ligand binding sites. Structure-based virtual screening was used to predict small drug-like molecules that may bind to those sites. These findings could help in studies aimed at obtaining novel insight into PB2 functions as well as provide a starting point for further in vitro investigations of replication inhibitors that are effective in a variety of hosts and lack the potential of inducing influenza antiviral resistance.

2. Results and discussion

2.1. PB2 sequence conservation

12,459 PB2 sequences were obtained from the NCBI influenza virus resource database. This included 31% from human, 16% from swine and 50% from avian hosts. 702 sequences remained after removing redundant sequences at 98.5% identity, indicating that a large proportion of PB2 sequences deposited are highly similar which would reflect bias upon conservation scoring (Valdar, 2002). The conservation scores calculated from the multiple sequence alignment of the non-redundant sequences shows that there is a high level of amino acid conservation throughout the entire protein sequence. The scores ranged from 0.789 (lowest) to 1.0 (highest) and the majority of amino acids had a score between 0.95 and 1.0 as shown in Fig. 1.

For display purposes the conservation scores were re-scaled and mapped on to the PB2 structure (Fig. 2). Overall, the key functional regions of PB2 were found to be highly conserved and consisted of several residues scoring 0.95 or above. This includes the N-terminal residues 1–37 which form three short α-helices comprising the PB1 binding interface required for effective polymerase activity (Sugiyama et al., 2009). The mGTP cap binding domain (318–482) is also well conserved, albeit with moderately conserved residues at position 339, 340, 453 and 456. Substitution of Lys339 to Thr339 of certain subtypes has been found to prevent binding of the phosphate group of mGTP capped mRNA, reducing RNA synthesis, and thereby regulating PB2 activity (Liu et al., 2013). Val414, Arg415 and Gly416 are highly conserved and are required for PB2-acetyl-CoA interaction to maintain transcription activity (Hatakeyama et al., 2014). The 424-loop region is

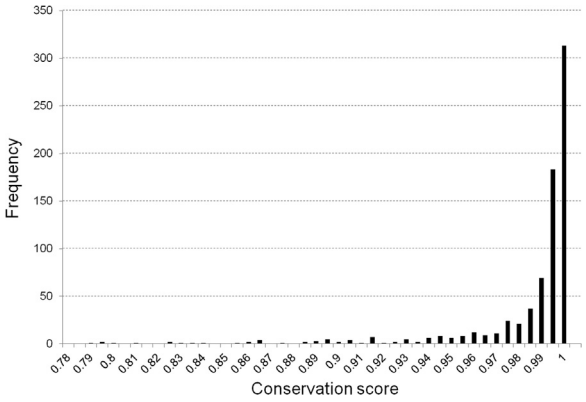

Fig. 1. Frequency distribution of PB2 amino acid conservation scores obtained after alignment of 702 non-redundant influenza A sequences from mainly human, avian and swine hosts using the Valdar scoring formula.

suggested to have an allosteric role in regulating PB1 activity, whilst other conserved residues are expected to contribute to the domains structurally distinct fold which allows formation of intermolecular contacts specific for mGTP cap binding activity (Guilligay et al., 2008). The 1–269 and 580–683 segments which are reported to be capable of binding the nucleoprotein (NP) (Poole et al., 2004), also consist of long stretches of conserved residues such as Ser592-Thr612. A total of 42 amino acids were found to be 100% conserved and could therefore be the most resistant to change due to evolutionary adaption of the virus. This includes Leu744 located on a surface exposed loop region and Gly693, which we suggest to be key residues in the NLS region due to their high conservation, enabling PB2 nuclear entry from the cytoplasm via binding importin-α. Other highly conserved regions with unassigned functions identified in this work may be of interest with regards to antiviral drug discovery.

An intermediate level of conservation for the host specific residue at position 627 was reflected in the alignment with a conservation score of 0.885. Due to the majority of sequences being from avian hosts, glutamic acid was the prevalent residue based on the consensus sequence. Whilst a range of amino acid residues can be tolerated at the 627 position shown by mutagenesis (Chin et al., 2014), the E627K mutation is well known for determining virulence by increasing polymerase activity and replication in mammals. This prime example of host adaptation is thought to be due to glutamic acid being able to bind the avian version of the host cell factor ANP32A; whereas substitution to lysine allows the polymerase to bind to the mammalian version of this host factor (Long et al., 2016; Moncorgé et al., 2010). However, some avian viruses carrying the E627 variant can efficiently replicate in mammalian cells due to compensatory mutations found in the PB1 protein of H5N1 strains (Xu et al., 2012). A mutation study of the 627 domain has also identified specific conserved residues to be essential for general PB2 activity (Arg597, Pro620, Phe621, Arg646 and Arg650), as well as non-essential residues such as Pro625, Pro626 and Gln628 which are also highly conserved (Kirui et al., 2014). Furthermore, the positive charge of the highly conserved Arg630 (in the presence of NP R150), or Lys627 promotes PB2-NP interaction, which is essential for the ribonucleoprotein complex to provide structural maintenance and regulate viral transcription (Labadie et al., 2007; Ng et al., 2012).

Low (scores below 0.85) or moderate conservation was identified mainly at single amino acid positions such as 64, 107, 147, 271, 292, 453, 483, 559, 588, 590, 591, 613, 661 and 676. The lowest conservation score was 0.789 at position 147. The residues at these positions are all located on the exterior surface of the protein (Fig. 2(b)), which is consistent with the finding that surface residues evolve faster than those in the protein core (Warren et al., 2013). Adaptive mutations to Ala271, Arg591, and Ser590 have been found to enhance polymerase