Antiviral Research 112 (2014) 120–131

Contents lists available at ScienceDirect

Antiviral Research

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

Global geno-proteomic analysis reveals cross-continental sequence conservation and druggable sites among influenza virus polymerases

Mustafeez Mujtaba Babar, Najam-us-Sahar Sadaf Zaidi *, Muhammad Tahir

Atta-ur-Rahman School of Applied Biosciences, National University of Sciences and Technology, H-12, Kashmir Highway, Islamabad 44000, Pakistan

ARTICLE INFO

Article history:
Received 30 June 2014
Revised 23 October 2014
Accepted 24 October 2014
Available online 31 October 2014

Keywords:
Influenza virus
Influenza polymerases
Drug targets
Geotemporal sequence conservation
Influenza protein druggability

ABSTRACT

Influenza virus is one of the major causes of mortality and morbidity associated with respiratory diseases. The high rate of mutation in the viral proteome provides it with the ability to survive in a variety of host species. This property helps it in maintaining and developing its pathogenicity, transmission and drug resistance. Alternate drug targets, particularly the internal proteins, can potentially be exploited for addressing the resistance issues. In the current analysis, the degree of conservation of influenza virus polymerases has been studied as one of the essential elements for establishing its candidature as a potential target of antiviral therapy. We analyzed more than 130,000 nucleotide and amino acid sequences by classifying them on the basis of continental presence of host organisms. Computational analyses including genetic polymorphism study, mutation pattern determination, molecular evolution and geophylogenetic analysis were performed to establish the high degree of conservation among the sequences. These studies lead to establishing the polymerases, in particular PB1, as highly conserved proteins. Moreover, we mapped the conservation percentage on the tertiary structures of proteins to identify the conserved, druggable sites. The research study, hence, revealed that the influenza virus polymerases are highly conserved (95–99%) proteins with a very slow mutation rate. Potential drug binding sites on various polymerases have also been reported. A scheme for drug target candidate development that can be employed to rapidly mutating proteins has been presented. Moreover, the research output can help in designing new therapeutic molecules against the identified targets.

1. Introduction

Influenza virus is one of the most important viral threats capable of crossing the species and geographical barriers. The virus is known to infect all warm blooded animals, both birds and mammals. The wide range of hosts indicates the ability of the virus to thrive in varying physiological, environmental and climatic conditions leading to significant mortality and morbidity in host organisms. In humans alone, the seasonal influenza virus infections have been associated with half a million to five million hospitalizations per annum worldwide (Maltezou, 2008; Thompson et al., 2004). Moreover, the 1918, 1957, 1968 and 2009 influenza pandemics have collectively recorded over 50 million deaths among infected individuals (Dawood et al., 2012; Guan et al., 2010; Morens et al., 2010). In addition, the disease and death in poultry and cattle lead to significant monetary losses across the globe (Pongcharoensuk et al., 2012). To address the global human and monetary losses, a number of prophylactic and therapeutic interventions have been developed. However, the high mutation rate helps the virus to resist these chemical and immunological measures. These mutations are attributed, mainly, to the error prone RNA genome of the virus. The influenza virus genome is composed of eight segments encoding 12 different proteins. The first three gene segments translate into 5 proteins which are collectively referred to as polymerase complex. The remaining five genes encode for the two surface glycoproteins (hemagglutinin and neuraminidase), matrix, nucleoprotein and the nonstructural proteins (Tsai and Chen, 2011). The high rate of genetic polymorphism and the adaptability of the virus to survive under varying physiological and geographical regions has been associated with the development of resistance against the currently available drugs (van der Vries, 2014). The focus, therefore, has to be on the identification and targeting of conserved functional molecular machinery of the virus for the drug design and development process.

Among the influenza proteins, the genetic conservation of influenza virus polymerases is very well established and has been

* Corresponding author at: Department of Industrial Biotechnology, Atta-ur-Rahman School of Applied Biosciences, National University of Sciences and Technology, H-12, Kashmir Highway, Islamabad 44000, Pakistan. Tel.: +92 51 90856132.
E-mail address: zaidi.sahar@gmail.com (N.S.S. Zaidi).

http://dx.doi.org/10.1016/j.antiviral.2014.10.013
0166-3542/

studied by a number of research groups (Chu et al., 2012; Warren et al., 2013; Wise et al., 2011). The 3P polymerase complex of the virus is encoded by segment 1 (Basic Polymerase 2 or PB2), segment 2 (Basic Polymerase 1 or PB1) and segment 3 (Acid Polymerase or PA). PB1 and PA are polycistronic genes both encoding for 2 proteins each. These polymerases have important regulatory and functional contribution in viral replication process. PB2 contributes to cap binding activity, nuclear localization signaling and nucleocytoplasmic trafficking (Poole et al., 2004). PB1 gene encodes for PB1 and PB1-F2. Both the N and C terminal domains of PB1 are highly conserved and possess the capability to bind to the PA and PB2 polymerase sub units respectively (Ghanem et al., 2007). PB1-F2, on the other hand is a small proapoptotic protein. The protein is involved in increasing the intensity of viral pathogenicity by decreasing viral clearance, thereby, increasing the viral titer (Chen et al., 2001). Similar to PB1, segment 3 of the genome encodes two proteins, PA and PA-X. The function of PA is not clear but it has been indicated to serve as a serine-protease (Hara et al., 2001). PA-X, encoded by the X-ORF of segment 3, is involved in suppressing host immune responses by inhibiting the inflammatory, apoptotic and immune signaling pathways (Muramoto et al., 2013). Hence, influenza polymerase complex collectively plays a pivotal role in establishing and mounting viral pathogenicity (Bradel-Tretheway et al., 2011). Inhibiting or blocking the function of either of these proteins can lead to effectively decreasing the viral titer in the host.

A number of studies have established the sequence conservation of the polymerases across various hosts of different geographical regions. Variations in temperature, humidity and other climatic parameters have been known to affect the viral transmission process (Lowen and Steel, 2014; Xiao et al., 2013). Moreover, the immunogenetic profiles of host organisms have also been linked to play a significant role in altering the viral survival, host-to-host transmission dynamics and its response to different drug molecules (Baigent and McCauley, 2003; Parra and Suarez-Kurtz, 2014). Establishment of the viral genetic conservation patterns across larger geographical units, however, needs to be made. Continent-wise genomic and proteomic sequence analysis would help to substantiate the development of universal drug candidates against the molecular regions that are conserved across the hosts belonging to various geographical and climatic conditions. Localization and migration patterns of most influenza virus hosts, with the exception of flying birds, are highly conserved within continental boundaries. The involvement of migratory birds, moreover, in inter-continental transfer of virus has also been a subject of discussion (Krauss et al., 2007; Lam et al., 2012; Pearce et al., 2010). Genetic and protein conservation of viral genome, therefore, has to be established at continental level while devising cross-host, cross-continent effective antiviral strategies.

The current study investigated the candidature of influenza polymerases as potential drug targets. We systematically analyzed over 137,000 nucleotide and amino acid sequences to understand conservation patterns of the proteins. The sequences were categorized on continental basis and meta-analysis was performed to identify the conserved regions of viral polymerases from different biogeographical regions of the world. We also, quantified the mutation rates and patterns of different polymerases to predict the possibility of mutations in viral genome in response to any antiviral agent. The degree of conservation of amino acids was also mapped on the tertiary structures of the proteins to identify novel drug binding pockets of polymerases. The scheme followed can help in designing interventional agents targeting the universally conserved regions. Moreover, the workflow can be adapted for designing specific inhibitors against various infectious and metabolic disorders based on the geogenomics approach.

2. Methods

2.1. Data collection and classification

The study comprises of a meta-analysis of gene and protein sequences of influenza virus polymerases. The total number of gene sequences (approx. 60,000) and protein sequences (approx. 78,000) selected for the analysis have been presented in Table 1. The sequences belonging to all host types and strains were grouped, thereafter, on the basis of continental diversity into six classes i.e., Africa, Asia, Europe, North America, Oceania and South America. Most of the influenza virus hosts are largely localized in a particular continent and their migration patterns are largely conserved within a continental boundary. The classification of sequences in this manner assisted in organization of data and determination of genetic diversity among hosts of varied geographical regions.

Gene segment sequences were obtained from Influenza Research Database (IRD) (Squires et al., 2012). Only complete segments of all the polymerases (Basic Polymerase 2 (PB2), Basic Polymerase 1 (PB1) and Acid Polymerase (PA)) were selected. The retrieved sequences contained complete coding regions. No advanced options including flu season, laboratory strains or date of collection were checked during the data collection process on the IRD. Similarly, amino acid sequences of all proteins (PB2, PB1, PB1-F2, PA and PA-X) encoded by the three influenza A virus polymerase genes were collected. Similar search parameters were employed for protein sequence collection from IRD, as those used for retrieving polymerase gene sequences.

2.2. Sequence conservation analysis

In order to analyze the sequence conservation patterns, collected samples were subjected to multiple sequence alignment (MSA) and position-wise polymorphism studies using the IRD. Multiple Sequence Alignment (MSA) was performed using MUSCLE

Table 1
Number of gene segment and protein sequences evaluated for the polymerases.

Geographic grouping	Number of gene sequences evaluated			Number of amino acid sequences evaluated				
	PB2	PB1	PA	PB2	PB1	PB1-F2	PA	PA-X
Africa	246	297	210	210	255	291	308	58
Asia	5362	5512	5324	5324	5483	3868	5639	863
Europe	2095	2137	2068	2068	2114	1034	2202	237
North America	9928	10465	9592	9592	9966	8859	10132	2563
Oceania	1228	1235	1185	1185	1186	984	1193	38
South America	600	602	596	596	598	432	597	390
	19,459	20,248	19,905	18,975	19,602	15,468	20,071	4149
	Total: 59,612			Total: 78,625				