World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

STRUCTURAL AND FUNCTIONAL INSIGHTS INTO THE REPLICASE/TRANSCRIPTASE SYSTEMS OF THE DEADLY BAT-BORNE HUMAN VIRUSES: MARBURG, EBOLA, AND NIPAH

Peramachi Palanivelu *

Department of Molecular Microbiology, School of Biotechnology, Madurai Kamaraj University, Madurai – 625 021, India.
* Corresponding Author

World Journal of Advanced Research and Reviews, 2026, 31(02), 1171–1190

Publication history: Received on 12 June 2026; revised on 18 August 2026; accepted on 20 August 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.31.2.2181

ABSTRACT

Replication and transcription of genomes in the human pathogenic viruses, viz. Marburg, Ebola and Nipah are accomplished by a virally encoded RNA-dependent RNA polymerase (RdRp). The enzyme performs this dual function with a polymerase cofactor, VP35, in Marburg and Ebola viruses and a phosphoprotein, P, in Nipah and its related viruses. As RdRp is indispensable for the viral lifecycle, it is one of the most attractive molecular targets for the development of broad-spectrum antiviral therapeutics to control the spread of these viruses. However, until now, little is known about the RdRp of these viruses, especially with respect to their active sites, polymerase catalytic centre and its catalytic mechanism, which are considered the most promising molecular targets for drug development initiatives. In this study, multiple sequence alignment (MSA) was integrated with structural information from X-ray crystallography, cryo-electron microscopy (cryo-EM) and site-directed mutagenesis (SDM) studies to identify the putative catalytic and metal-binding sites of the RdRp from Marburg, Ebola and Nipah viruses. MSA analysis suggests that the RdRp from these viruses uses an -LA/VG- as the template-binding pair, an invariant K, as the proton abstractor which initiates the catalysis and an invariant R at -4 from the catalytic K, as the nucleotide discriminator, all of which are highly or completely conserved. Besides, the Marburg and Ebola viruses contain an additional second putative polymerase catalytic core within the N-terminal domain (NTD), which is absent in the Nipah virus polymerase. The identified catalytic residues exhibit strong structural similarity to the conserved polymerase catalytic site architectures already reported in numerous RNA- and DNA-dependent polymerases, supporting a common catalytic framework. The catalytic magnesium-binding sites are also completely conserved among these viruses and consist of two invariant aspartate-containing motifs, -G**D**N- and -T**D**L-. This arrangement differs slightly from that what is reported for several other non-segmented negative-sense RNA viruses, including influenza, rabies, mumps and measles, which typically use -G**D**N- and -G/S**D**D- motifs for catalytic magnesium-binding. Furthermore, the RdRp domains of all three viruses contain two putative zinc-binding motifs (ZBMs), suggesting an important structural role in polymerase stability and function. Interestingly, the RdRp of Marburg and Ebola viruses, members of the *Filoviridae* family, possess an additional putative HNH-endonuclease motif(s) that is absent in the Nipah virus polymerase, a member of the *Paramyxoviridae* family. Collectively, these findings reveal conserved structural and catalytic features of the polymerases from these bat-borne human pathogens, and provide new insights into their evolutionary relationships, and also identify potential molecular targets for rational antiviral drug design.

Keywords: Bat-borne human viruses, Marburg virus, Ebola virus, Nipah virus, RNA-dependent RNA polymerase, HNH-endonuclease, Polymerase metal-binding sites, Polymerase catalytic site.

1 INTRODUCTION

Viruses are ubiquitous in nature and infect organisms across all kingdoms of life. Based on their genetic material, they are broadly classified as RNA and DNA viruses. Both groups cause severe infectious viral diseases of both endemic and pandemic proportions, highlighting the need for a deeper understanding of their pathogenic mechanisms and replication strategies in their respective host cells in order to develop effective antiviral therapeutics [1]. However, the majority of emerging and reemerging human and animal diseases of endemic and pandemic potential are caused by the RNA viruses such as SARS-CoVs, Marburg, Ebola, Nipah, Rabies, etc. [2]. For example, the ongoing Ebola outbreak, caused by the Ebola (Bundibugyo strain) in the Democratic Republic of the Congo, for which no licensed vaccine or approved specific therapeutic is currently available, has resulted in 2378 deaths and about 5000 confirmed cases as of 19th August 2026 [Reuters]. The DNA and RNA polymerases which involve in the replication of their genomes are collectively referred to as DNA and RNA replicases, respectively. Their primary function is to ensure accurate replication of their respective genomes so that genetic information is faithfully transmitted to the next generation. Analyses of the active sites of hundreds of DNA and RNA polymerases from viruses to humans, have revealed a remarkable degree of structural and functional conservation. It is interesting to note that in spite of their evolutionary diversity, almost all these DNA/RNA polymerases adopt the same overall 3D structure with a characteristic cupped 'right hand' architecture comprising three conserved subdomains, viz. palm, fingers, and thumb, suggesting the chemical basis for designing the active sites using the same or functionally similar amino acids to perform the polymerization reactions in all organisms. Each of these subdomains performs a distinct, yet highly coordinated function during template recognition, nucleotide selection, and catalysis during the polymerization process in all organisms [3, 4]. Genome replication by DNA and RNA polymerases is generally associated with two fundamental functions: sequential addition of nucleotides on a template following Watson-Crick base pairing and proofreading; the latter ensures replication fidelity by removing misincorporated nucleotides [3-7]. The RdRp of these highly pathogenic bat-borne human viruses is unique in that the single enzyme performs dual function, that is, it replicates the viral genome and also transcribes the viral genes and thus, making the enzyme central to the viral lifecycle. Its indispensable role in viral transcription and replication make the enzyme as one of the highly attractive targets for the development of broad-spectrum antiviral drugs.

All three bat-borne viruses possess a negative-strand RNA genome. The negative-strand RNA viruses are broadly classified into segmented and non-segmented viruses. Examples of families of non-segmented, negative-strand RNA viruses (nsNSVs) with a single-stranded RNA genome belong to the order *Mononegavirales* and include the following families: *Filoviridae* (e.g., Marburg virus (MARV) and Ebola virus (EBOV), *Paramyxoviridae* (e.g., Nipah virus), *Pneumoviridae* (e.g., respiratory syncytial virus), *Rhabdoviridae* (e.g., rabies virus), *Bornaviridae* (e.g., Borna viral diseases, *Nyamiviridae* (e.g., nyaviruses that infect birds and reptiles), *Mymonaviridae* (e.g., mycoviruses that infect fungi), etc. The order *Mononegavirales* includes many of the important human pathogenic viruses, such as rabies (RABV), MARV, EBOV, Nipah (NiV), measles (MeV), mumps (MuV), etc. The MARV and EBOV are enveloped, bat-borne nsNSVs, capable of causing severe human diseases with potentially fatal outcomes. They cause filovirus disease characterized by severe haemorrhagic fever with up to 90% fatalities [8]. However, the other bat-borne virus, the Nipah virus, is also enveloped nsNSV from the *Henipavirus* genus, and causes lethal encephalitis with a mortality rate of approximately 90%. (However, the SARS-CoVs, MERS-CoV and other human CoVs are also enveloped viruses and possess a non-segmented, but *positive-strand* RNA genome). The classification of these bat-borne viruses, their natural reservoir(s) and their intermediary hosts are discussed in the previous publication by the author [7].

The MARV and EBOV which belong to the *Filoviridae* family, consists of two genera: EBOV with five distinct species, including the *Zaire Ebolavirus* (EBOV) and the *Marburgvirus* (MARV) with a single species. The RNA genomes of EBOV and MARV are approximately 19 kb in length and encode seven structural proteins (Fig. 1). The proofreading (PR) function in these three bat-borne viruses was found to be associated with the PRNTase domain and a detailed analysis of the PR domain has already been reported [7]. In this communication, a structural and functional analysis of the replicase/transcriptase systems of these bat-borne human pathogenic viruses is reported.

1.1 Genome organization of Marburg, Ebola and Nipah Viruses

As mentioned above, the MARV and EBOV genomes are approximately 19 kb in size and encode seven viral proteins in tandem. They include: nucleoprotein (N) and other viral proteins such as VP35, VP40, glycoprotein (GP), VP30, VP24, and a large protein (L) (Fig. 1A) [9, 10]. They produce monocistronic mRNAs, corresponding to each of the viral genes with the 'gene-start (GS) and gene-end (GE)' signals on the viral RNA.

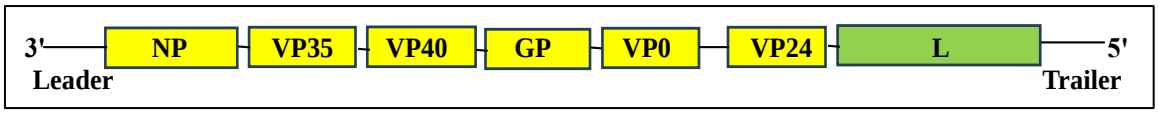


Figure 1A: Schematic diagram showing the organization of viral genes in EBOV and MARV genomes (~19 kb); (Zaire EBOV-18,957 nts; Lake Victoria MARV-19,111 nts). ▁, intergenic regions (IRs)

Ebola and Marburg viral genes are arranged in a linear order (3' -NP-VP35-VP40-GP-VP30-VP24-L-5 ') (Fig. 1A). The genes are flanked by conserved GS and GE signals that either overlap or are separated by short IRs, generating each mRNA with a 5'-cap and 3'-polyadenylated tail. However, during replication the polymerase synthesizes a full-length complementary antigenome as an intermediate for genomic RNA production regardless of the transcriptional signals. The cofactor VP35 is a multifunctional protein which is not only required for suppression of the host's innate immune response, but also cooperates with the L protein to perform viral genome replication and mRNA transcription [10, 11]. NP, VP35, and L, were found to be essential and sufficient for replication. In contrast to MARV, the EBOV-specific transcription is dependent on the presence of a fourth nucleocapsid protein, VP30 [12]. Located at the ends of the genomes are the 3' leader (Le(−)) and the 5' trailer (Tr(−) [10].

On the other hand, the NiV genome consists of six transcriptional units, encoding Nucleoprotein (N), Phosphoprotein (P), Matrix protein (M), Fusion protein (F), Glycoprotein (G), and the large protein (L). These genes are transcribed in a sequentially attenuated order with the consensus 'gene-start (GS) and gene-end (GE)' sequences generating each mRNA with a 5'-cap and 3'-polyadenylated tail similar to MARV and EBOV [13]. Located at the ends of the NiV genome are the 3' leader (Le(−)) and the 5' trailer (Tr(−) as in MARV and EBOV genomes. However, during replication, the polymerase synthesizes a full-length complementary antigenome as an intermediate for viral genomic RNA production regardless of the transcriptional signal sequences as in MARV and EBOV. It is interesting to note that the 'P' ORF also encodes three more non-structural proteins, viz. C, V and W, which play key roles in NiV's pathogenicity [14] (Fig 1B).

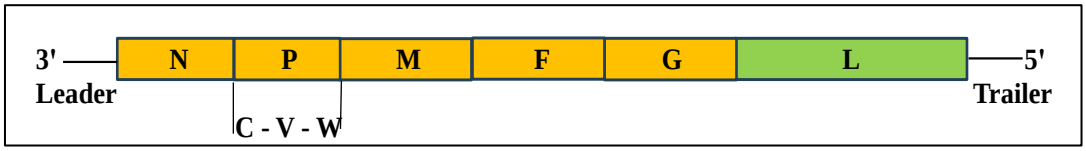


Figure 1B: Schematic diagram showing the organization of the genes in the NiV genome (~18 kb; 18, 246 nts).

The RNA of the P gene is edited to generate the V and W proteins, when additional guanines are inserted at the RNA editing site due to the stuttering of the RNA polymerase during transcription. Thus, the P, V, and W proteins share a common N-terminal domain, but each has a unique C-terminal domain. On the other hand, the C protein is encoded in an alternative reading frame of the P gene [15].

1.2 Organization of the L protein in Marburg, Ebola and Nipah Viruses

The L protein is extensively studied and has a molecular mass of ~250 kDa, and consists of at least seven discernible domains, viz. a N-terminal domain (NTD), an RdRp domain, a Guanosine diphosphate Polyribonucleotidyl Transferase (PRNTase) domain, a Polymerase proofreading exonuclease (PR Exo) domain, a Connector domain (CD), a Methyltransferase (MTase) domain, and a C-terminal domain (CTD). These enzymes and domains are organized contiguously on the L protein as shown in Fig. 2. In MARV, EBOV and NiV, both transcription and replication are mainly accomplished by the same enzyme, i.e., by an RdRp, (EC 2.7.7.48). Other enzymes and functional domains of the L protein also play a key role in the transcription and replication. For example, the PRNTase performs capping of the mRNAs and the MTase is a bifunctional enzyme and exhibits dual specificity, i.e., it methylates the ribose at 2′-O position on the first nucleotide of the transcript and then methylates the capped 5'-guanylate at N-7 position. The CD and the CTD, both with nonenzymatic functions, flank the MTase. It is found that the PRNTase domain also harbours the proofreading 3'→5'exonuclease function [7]. In MARV and EBOV, the VP35 acts as a non-catalytic cofactor for the RdRp whereas a phosphoprotein, P acts as a non-catalytic cofactor for the RdRp of NiV. Interestingly, all these essential enzymes are considered for the development of potential antiviral drugs for effective control of these viruses, which is further elaborated in the following sections.

The replication and transcription of viral genomes are mediated by a polymerase complex consisting of two proteins: L and its cofactor VP35/P [16,17]. The polymerase cofactor VP35 acts as a trimer in MARV and EBOV RdRps, and the cofactor P (~80 kDa), a phosphoprotein, acts as a tetramer in NiV RdRp [18]. The entire NTD, up to amino acid ~380 is required for strong VP35–L/ P-L binding.

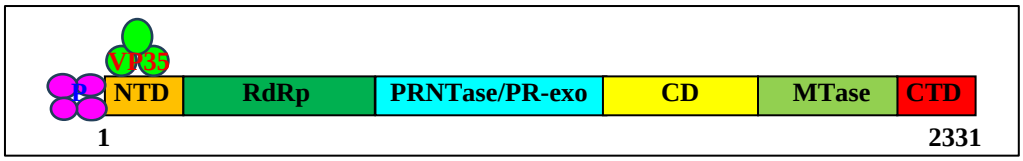


Figure 2: Organization of the enzymes and domains of the L protein in Marburg (2331), Ebola (2212) and Nipah (2244) viruses.

In the first phase of their lifecycle, the viruses enter the host cell via receptor binding and membrane fusion (Both MARV and EBOV share an essential intracellular endosomal receptor called Niemann-Pick C1 (NPC1), and in NiV, the G protein binds to the cell-surface receptor, viz. ephrin-B2 or -B3). The viral ribonucleoprotein complex, (consisting of the genomic RNA encapsidated by the nucleoprotein), is released into the cytoplasm of the host cells. After uncoating, the viral RNA acts as a template for synthesis of viral mRNAs, which is followed by protein synthesis using the protein synthetic machinery of the host cells. In nsNS RNA viruses, the RNA genome is encapsulated by NP and further forms a ribonucleoprotein (RNP) complex with L protein and its RdRp cofactor VP35/P (and additionally with the VP30 in EBOV). In the transcription mode, the polymerase initiates at the 3′ genome promoter and produces capped, methylated, and polyadenylated monocistronic mRNAs by recognizing 'gene-start' and 'gene-end' signals that flank each gene. In replication mode, the same enzyme bypasses the regulatory signals to synthesize a full-length (+) strand antigenome which serves as the template for the synthesis of new (−) sense strand viral genomic RNAs. Assembly of viral particles begins with the formation of RNPs which accumulate in the perinuclear region and are transported by the microtubules to the budding sites on the plasma membrane.

The mode of replication and transcription is very similar for these viruses. For example, the transcription and replication are accomplished by the promoter at the 3'-leader region of the genome and antigenome. (It is interesting to note that the nsNS RNA genomes possess complementary nucleotides at their 3'- and 5'-terminals. And because of this complementarity, the promoters at the genome and the antigenome are almost identical to each other. Therefore, the transcription and replication start from the same promoter element at the 3'-leader of the genomic and the antigenomic RNAs). RNA synthesis by multiprotein complexes of EBOV, human influenza B, RSV and monomeric enzymes of hepatitis C and Zika viruses require a 5'-phosphorylated primer with a minimum primer length between two and three nucleotides. During the replication mode, the RdRp with NP, VP35/P bound to the viral RNA becomes super processive, and ignores the regulatory sequences and replicates the entire viral genome without any pause. Thus, the viral RNA replication is a two-step process, i.e., in the first step, a full-length copy of the genome with positive polarity is synthesized, which is also known as the antigenomic RNA. The antigenomic RNA, that is the plus-strand intermediate RNA, is concurrently encapsidated by the viral nucleocapsid protein N. The encapsidated antigenome now serves as a template for the synthesis of viral genomic RNA, i.e., the negative-strand RNA [7]. It has been shown that the replication and transcription of viral genomes are mediated by a polymerase complex essentially consisting of two proteins: L protein and its cofactor VP35 in MARV and EBOV which is analogous to the P protein of other nsNSVs [17,18].

1.3 The viral RdRp as a potential drug target in Marburg, Ebola and Nipah viruses

Consistent with their shared genome structure, the nsNS RNA viruses have evolved similar ways to transcribe their genomes into mRNAs and then to replicate to produce new genomes. A single viral RdRp (of the L protein) performs both functions. The transcribed viral mRNAs are then capped, methylated by the other two enzymes (PRNTase and MTase) of the L protein. The poly-A tails are added to the mRNAs at their 3'-ends by a stuttering mechanism at a slippery stop-site present at the end of the viral genes. The capped and tailed mRNAs are then translated into viral proteins in the cytoplasm using the host cell machinery. In the transcription mode, the polymerase performs the sequential transcription of all the mRNAs using a 'termination-reinitiation' mechanism that responds to 'GS and GE' signals. Some polymerases could also disengage from the template at each gene junction, resulting in decreasing abundance of transcripts from the 3' to the 5' end of the genome. Once enough nucleocapsid proteins (N) are made from its mRNA, the viral RNA is now tightly encapsidated by the N and thus, the RdRp switches from transcription to the replication mode, which is tightly linked to the intracellular concentrations of N. As discussed elsewhere, in the replication mode, the polymerase replicates the entire viral genome without recognizing the transcription signals, and the replicated genome is not capped or polyadenylated. Thus, the RdRp of these RNA viruses plays a dual role, first as a transcriptase and then as a replicase (except in retroviruses) in animal and human cells. Therefore, RdRp from these viruses has been the main target for antiviral drug development to control their proliferation in animal and human cells.

2 MATERIALS AND METHODS

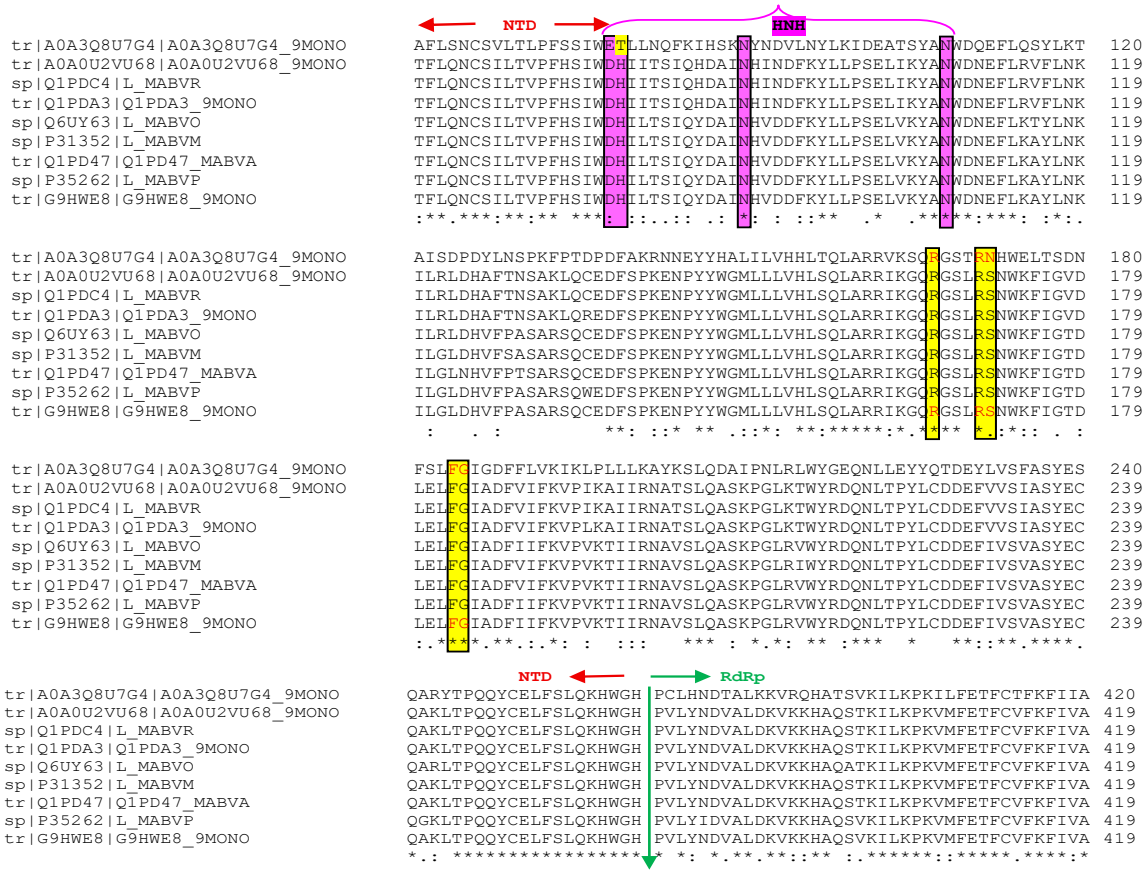
Full-length L protein sequences from the MARV, EBOV and NiV were obtained from the PubMed and Swiss-Prot databases. The advanced version of Clustal Omega was used for MSA analysis [19, 20]. The conserved motifs and sequence similarities identified by the bioinformatics analysis are integrated with the data already available from

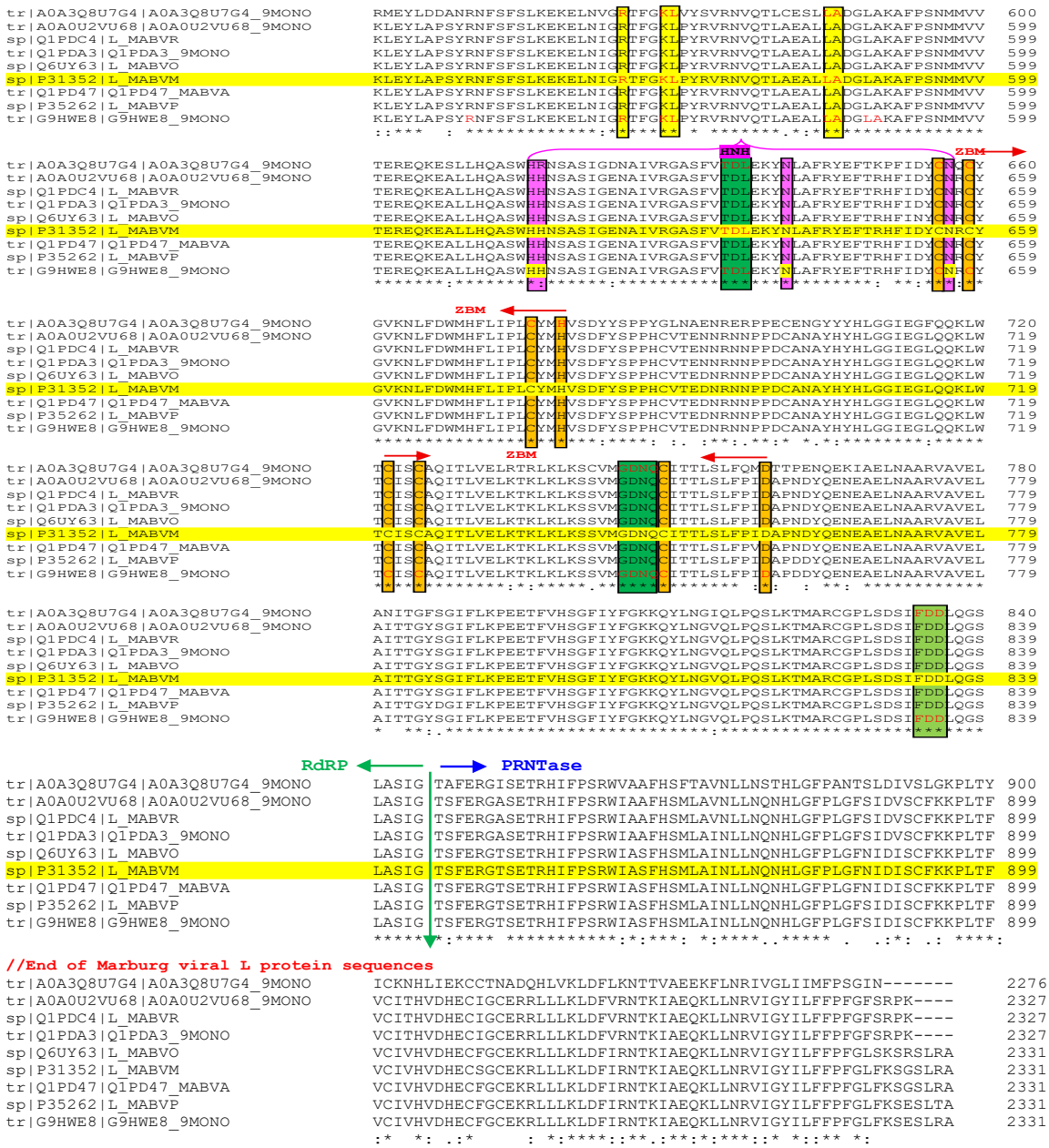
biochemical, SDM, cryo-EM, and X-ray crystallographic analyses on these enzymes to arrive at the potential polymerase active site amino acids in the RdRps of the MARV, EBOV and NiV.

3 RESULTS AND DISCUSSION

Figure 3 shows a part of the NTD and RdRp domains in the L protein of different strains of the MARV (only the region required for the discussion is shown here). The MARV (MARV/Mus) is a highly virulent variant of the MARV which was identified from a human case in 1980 in Kenya and is used as the standard strain and is highlighted in yellow. Besides, Musoke is a cornerstone strain for understanding filovirus infections in general, which is characterized by a severe, often fatal, haemorrhagic fever and multi-organ failure in humans and non-human primates [21]. The tentative region that harbours the RdRp domain is shown by arrows. The region is highly conserved with a large number of conserved peptides from the N-terminal to the end of the RdRp domain. The NTD harbours a putative -DH- type HNH-endonuclease motif except in *Dianlovirus menglaense* L protein (highlighted in magenta). Interestingly, a complete set of putative conserved amino acid residues constituting the polymerase catalytic core was identified by sequence similarity within the RdRp domain and is highlighted in yellow. This catalytic core identification is strongly supported by the data from RNA/DNA polymerases from viruses to humans (Table 1). Interestingly, all the strains of the MARV use the same completely conserved catalytic core, viz.- ^{561}NIGR-^{4}TFGK^{568}LPYRVRNVQTLAEALLADGLAK- as shown in Fig. 3. In most of the RNA/DNA polymerases the template-binding pair is -Y/F/V/LG-. However, an -LA- pair is identified here by sequence similarity and distance conservations. The metal-binding motifs, viz. -^{634}TDL- and -^{744}GDNQ- are completely conserved in all. These data are further confirmed by the cryo-EM map of the MARV polymerase [16]. The invariant Ds in these motifs are implicated in the catalytic Mg^{2+}-binding. (Interestingly, a tripeptide tandem repeat is observed within the polymerase catalytic core (marked by arrows) (Table 1). The NTD of the MARV also harbours a putative polymerase catalytic core region (highlighted in yellow). Additionally, a completely conserved -FDD -motif (highlighted in light green) is also found in the RdRp of all strains (Fig. 3). Other interesting features of the MARV RdRp are the presence of a putative -HH- type HNH-endonuclease motif (highlighted in magenta) and two putative Zinc-binding motifs (ZBMs) (highlighted in orange) [22]. The HNH-endonuclease motif and its relevance are discussed elsewhere.

CLUSTAL O (1.2.4) MSA of the NTD and RdRp domains from different strains of MARV and from various regions.





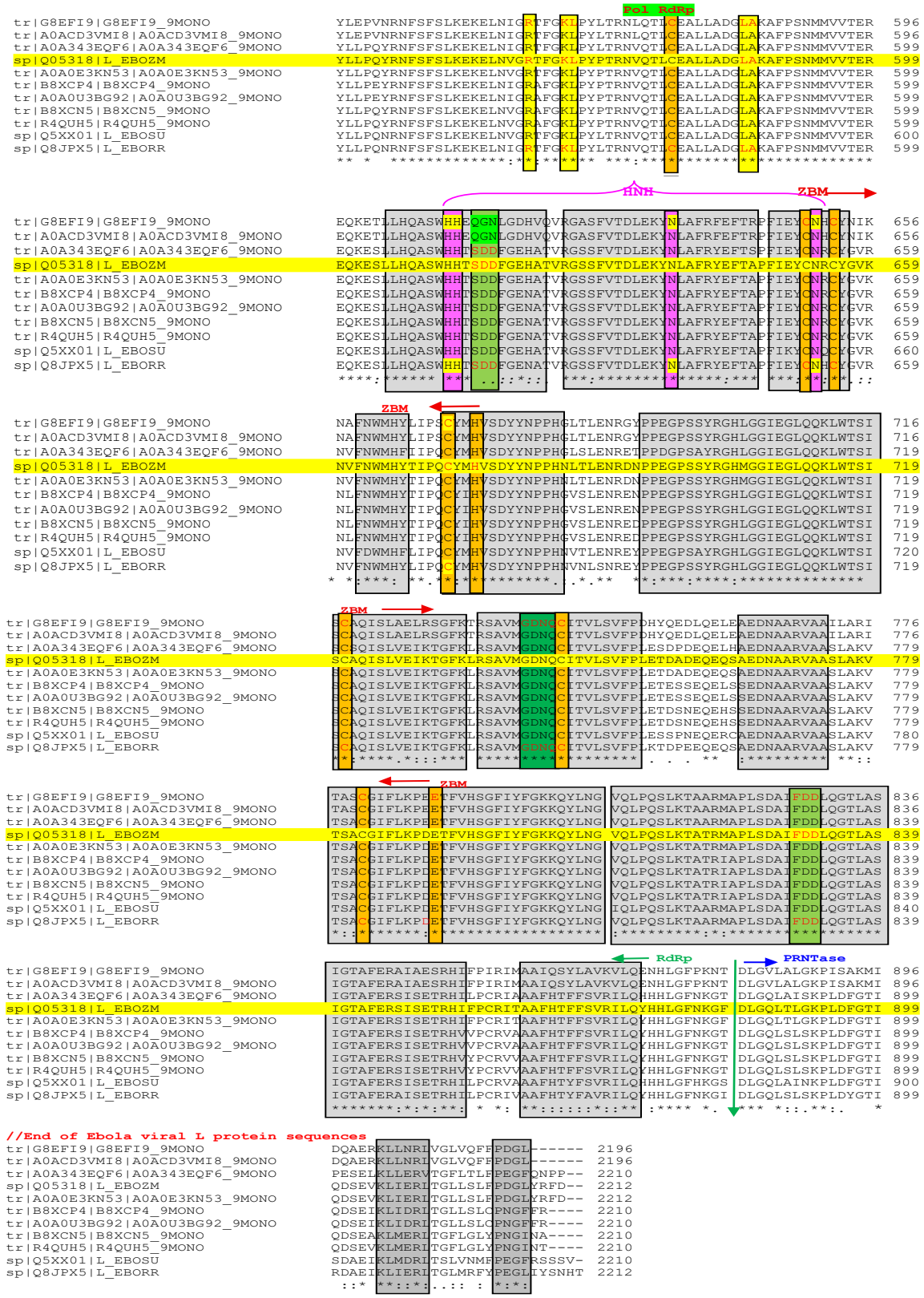
A0A3Q8U7G4_9MONO RNA-directed RNA polymerase L, *Dianlovirus menglaense*
A0A0U2VU68_9MONO RNA-directed RNA polymerase L, *Orthomarburgvirus marburgense*
Q1PDC4|L_MABVR RNA-directed RNA polymerase L, Lake Victoria Marburgvirus (strain Ravn-87)
Q1PDA3_9MONO RNA-directed RNA polymerase L, Lake Victoria Marburgvirus - DRC1999
Q6UY63|L_MABVO RNA-directed RNA polymerase L, Lake Victoria Marburgvirus, (strain Ozolin-75)
P31352|L_MABVM RNA-directed RNA polymerase L, Lake Victoria Marburgvirus (strain Musoke)
Q1PD47_MABVA RNA-directed RNA polymerase L, Lake Victoria Marburgvirus (strain Angola 2005)
P35262|L_MABVP RNA-directed RNA polymerase L, Lake Victoria Marburgvirus (strain Popp-67)
G9HWE8_9MONO RNA-directed RNA polymerase L, Lake Victoria Marburgvirus – Leiden

Figure 3: MSA of the NTD and RdRp domains from different strains of MARV and from various origins

Figure 4 shows a part of the NTD and RdRp domains from different strains of EBOV and from different geographical regions. EBOV (Zaire) is used as the standard and highlighted in yellow (only the region required for the discussion is shown here). The tentative region that harbours the RdRp domain is shown by arrow marks. The NTD is not highly conserved except for a region at the very beginning of the domain which shows a putative ZBM and a putative polymerase (Pol2) active site core (highlighted in yellow). However, the RdRp domain shows a large number of highly conserved peptides. Interestingly, a complete set of conserved polymerase catalytic core (Pol1) active site amino acids are identified in this region (highlighted in yellow). These identified catalytic core active site amino acids are in close agreement with a large number of RNA/DNA polymerases from viruses to humans (Table 1). Interestingly, all the strains of the EBOV use the same completely conserved amino acids in the catalytic core, viz.- ^{558}NVGR-^{4}TFGK^{565}LPYPTRNVQTLCEALLADGLAKAFP- as shown in Fig. 4. In most of the RNA/DNA polymerases the template-

binding pair is -Y/F/V/LG-. However, an -LA- pair is found here by sequence similarity as in MARV RdRp. The metal-binding motifs, viz. -^{631}TDL- and -^{741}GDNQ- are completely conserved in all. Muhlberger *et al*. and Schmidt *et al*. [12, 23] provided the first experimental evidence to show that the -GDNQ- motif is indeed essential for both EBOV genome replication and transcription. This was further corroborated by MSA of the RdRps from a large number of non-segmented, negative strand viruses. When this motif was mutated by substituting of the -DNQ- sequence to 3 Ala residues, they found that this motif was essential for both viral genome replication and transcription. Their finding was further corroborated by an additional SDM experiment where only the **D** in the -G**D**^{742}NQ- was substituted to an **Ala**^{742} and the EBOV **D**^{742}**→A** mutant enzyme lacked the ability to extend the primer in the presence of Mg^{2+} [24]. The EBOV also harbours an additional putative Pol 2 catalytic core region at the NTD (highlighted in yellow). The 3D structure showed that the RdRp domain of EBOV L folds into the canonical right-handed fingers–palm–thumb architecture observed in most of the RNA viral polymerase structures [17]. They also further confirmed that the D^{742}→A mutant did not produce any transcript. Additionally, as in MARV RdRp, an -FDD -motif is completely conserved (highlighted in light green) in all (Fig. 4). Recently, it was found that the MARV polymerase and the invariant Ds in these motifs are implicated in the catalytic Mg^{2+}-binding by cryo-EM mapping [16]. Interestingly, a tripeptide tandem repeat is observed within the polymerase catalytic core (marked by arrows) as in MARV. Other interesting features of the EBOV RdRp are the presence of a putative HNH- endonuclease motif (highlighted in magenta) and two putative zinc-binding motifs (ZBMs) (highlighted in orange). The HNH-endonuclease motif and its relevance is discussed elsewhere.

CLUSTAL O (1.2.4) MSA of the NTD and RdRp domains of different strains of EBOV and from different origins



G8EFI9_9MONO RNA-directed RNA polymerase L, *Cuevavirus lloviuense*
A0ACD3VMI8_9MONO RNA polymerase, *Cuevavirus lloviuense*
A0A343EQF6_9MONO RNA-directed RNA polymerase L, Bombali virus
Q05318|L_EBOZM RNA-directed RNA polymerase L, Zaire Ebolavirus (strain Mayinga-76)
A0A0E3KN53_9MONO RNA-directed RNA polymerase L, Ebola virus
B8XCP4_9MONO RNA-directed RNA polymerase L, Tai Forest ebolavirus
A0A0U3BG92_9MONO RNA-directed RNA polymerase L, Tai Forest ebolavirus
B8XCN5_9MONO RNA-directed RNA polymerase L, Bundibugyo virus
R4QUH5_9MONO RNA-directed RNA polymerase L, Bundibugyo virus
Q5XX01|L_EBOSU RNA-directed RNA polymerase L, Sudan ebolavirus (strain Human/Uganda/Gulu/2000)
Q8JPX5|L_EBORR RNA-directed RNA polymerase L, Reston ebolavirus (strain Reston-89)

Figure 4: MSA of the NTD and RdRp domains of different strains of EBOV and from different origins

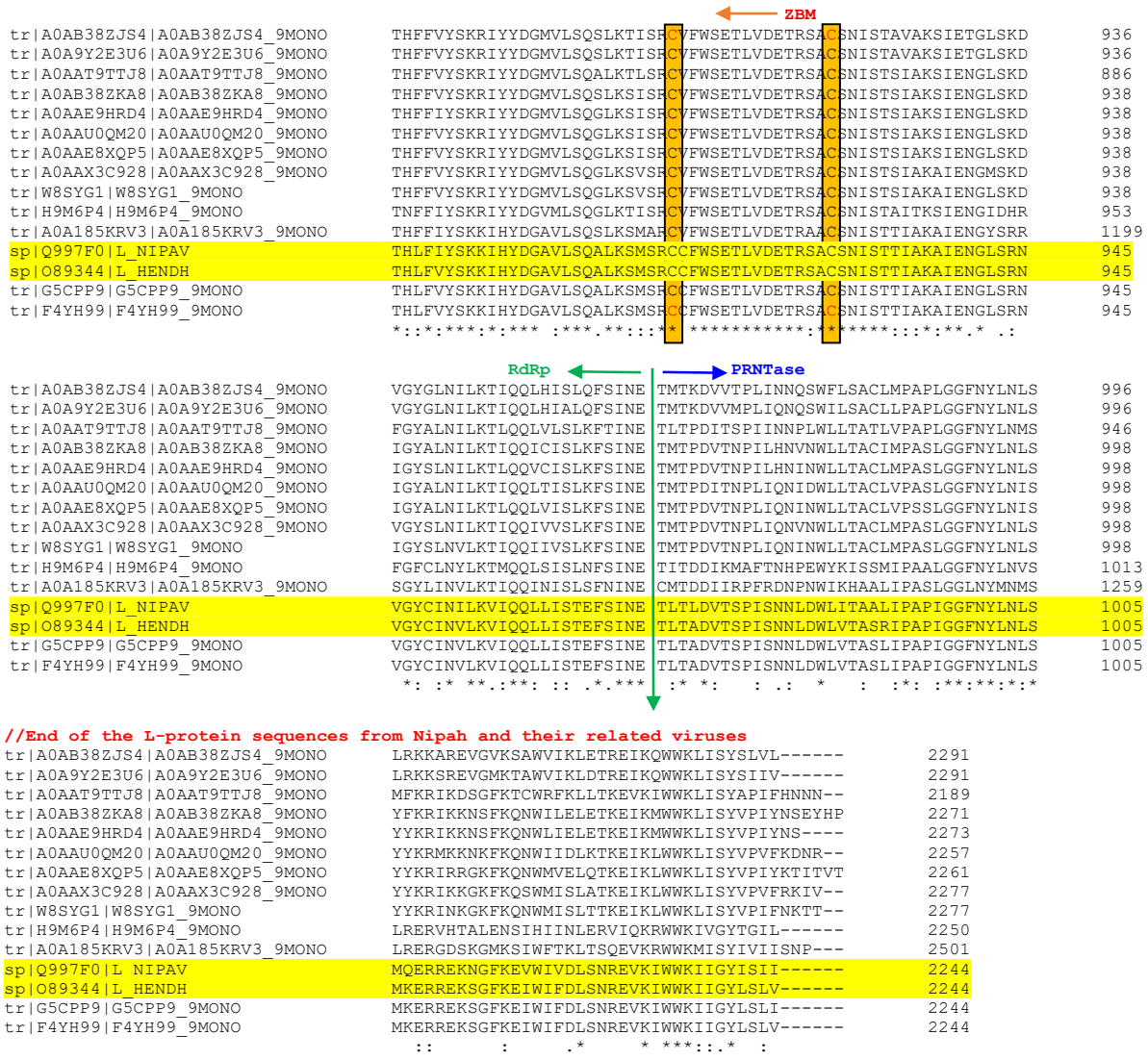
Figure 5 shows the MSA of a part of the NTD and RdRp domains of NiV and its related viruses. Nipah and its related species. The NiV and its related Hendra virus are used as the standards and highlighted in yellow (only the region required for the discussion is shown here). The tentative region that harbours the RdRp domain is shown by arrow marks. The N-terminal region is not highly conserved and shows numerous of gaps and a few highly conserved large peptides are found in this domain (data not shown). However, the RdRp domain shows high conservations and a large number of highly conserved residues. Interestingly, a complete set of conserved, putative catalytic core active site amino acids are identified in this region and are highlighted in yellow. This catalytic core is in close agreement with a large number of RNA/DNA polymerases from viruses to humans (Table 1). Interestingly, all the strains of the Nipah and their related viruses use the same set of completely conserved amino acids in the catalytic core, viz. -^{548}QAGR-^{4}LF^{553}AK^{555}MTYKMR^{561}ACQVIAEALIASGVGKYF- as shown in Fig. 5 (the amino acids highlighted grey are from cryo-EM data). The two completely conserved metal-binding motifs, viz. -^{721}TDL- and -^{830}GDNQ- are found in the RdRp of NiV species as reported for the other two bat-borne viruses, viz. MARV and EBOV. Wang *et al.* [25] have shown that the -^{830}GDN- motif is found to be in the active site region of the polymerase by cryo-EM analysis which is in agreement with the MSA analysis. Furthermore, they found that the six amino acids that line the channel, Lys^{474}, Arg^{504}, Arg^{542}, **Phe^{553}, Arg^{561}**, and Lys^{804}, are conserved among the EBOV, RSV, and NiV RdRp structures. They suggested that the five positively charged amino acids are likely to interact with the RNA backbone, whereas the sixth one, the **Phe^{553}**, could interact with RNA bases (guanine in EBOV and cytosine in RSV, respectively) through π-π interactions [25]. It is interesting to note that these two amino acids, **Phe^{553}, Arg^{561}** (highlighted in grey) are also found within the catalytic core of the NiV RdRp by MSA analysis (Fig. 5). The MSA analysis has also identified an additional DxD type metal-binding motif (highlighted in light green).

Recently, structural insights into the NiV RdRp complex have emerged at an unprecedented pace. It is interesting to note that Peng *et al*. [26] found by cryo-EM analysis that one of the critical amino acids identified in the catalytic core (R^{551}) by MSA analysis in the present study, is actually in the binding interface amino acid residues, viz. E^{291}, K^{542}, **R^{551}**, K^{724}, N^{833}, and K^{893} which are conserved in all the nsNS RNA viruses, suggesting the broad-spectrum inhibition of suramin (a synthetic molecule and a century-old medication primarily used to treat the early stage of African sleeping sickness and river blindness) against different nsNS RNA viruses could be by occluding the substrates from entering the active site. A recent report by Sala and Hellen [27] found that the incoming nucleotide is stabilized within the active site groove through interactions with **R^{551}** from the fingers' domain, the template C10 base, and a coordinated magnesium ion. Their findings further corroborate the MSA analysis result, in identifying **R^{551}** in the nucleotide discrimination site. In support of this conclusion, Hu *et al.* [28] found by cryo-EM mapping that the side chain of **R^{551}** interacts with the phosphates of the incoming nucleotide during RNA elongation, while the side chains of **D^{832}** of -G**D^{832}**NE- motif and **D^{722}** of the -T**D^{722}**L- motifs coordinate the active site metal ion. They also found that the **D^{832}→ A** mutant was inactive. These findings were further confirmed by Balıkçı *et al*. [29] where substitution of the catalytic active site Asp with Ala at position 832 (**D^{832}→A**), resulted in the absence of any RNA product, indicating the loss of polymerase activity. Interestingly, a tripeptide tandem repeat is observed within the polymerase catalytic core (marked by arrows) as in MARV and EBOV.

In addition to the above conserved motifs, two putative ZBMs are also found in this domain (highlighted in orange) which could play a role in maintaining the structural integrity of the RdRp. A marked difference between the MARV and EBOV RdRps is the absence of the HNH-endonuclease motif in the NiV RdRp (Fig. 4).

CLUSTAL O (1.2.4) MSA of RdRp domains of Nipah and its related viruses.


ZBM
tr|A0AB38ZJS4|A0AB38ZJS4_9MONO
tr|A0A9Y2E3U6|A0A9Y2E3U6_9MONO
tr|A0AAT9TTJ8|A0AAT9TTJ8_9MONO
tr|A0AB38ZKA8|A0AB38ZKA8_9MONO
tr|A0AAE9HRD4|A0AAE9HRD4_9MONO
tr|A0AAU0QM20|A0AAU0QM20_9MONO
tr|A0AAE8XQP5|A0AAE8XQP5_9MONO
tr|A0AAX3C928|A0AAX3C928_9MONO
tr|W8SYG1|W8SYG1_9MONO
tr|H9M6P4|H9M6P4_9MONO
tr|A0A185KRV3|A0A185KRV3_9MONO
sp|Q997F0|L_NIPAV
sp|O89344|L_HENDH
tr|G5CPP9|G5CPP9_9MONO
tr|F4YH99|F4YH99_9MONO
ZBM



A0AB38ZKA8_9MONO RNA-directed RNA polymerase L, Jingmen Crocidura shantungensis henipavirus
A0A9Y2E3U6_9MONO RNA-directed RNA polymerase L, Crocidura tanakae henipavirus
A0AAT9TTJ8_9MONO RNA-directed RNA polymerase L, Chodsigoa hypsibia henipavirus
A0AB38ZJS4_9MONO RNA-directed RNA polymerase L, Jingmen Crocidura shantungensis henipavirus
A0AAE9HRD4_9MONO RNA-directed RNA polymerase L, Melian virus
A0AAU0QM20_9MONO RNA-directed RNA polymerase L, Lechcodon virus
A0AAE8XQP5_9MONO RNA-directed RNA polymerase L, Wenzhou Apodemus agrarius henipavirus
W8SYG1_9MONO RNA-directed RNA polymerase L, Mojiang virus
A0AAX3C928_9MONO RNA-directed RNA polymerase L, Langya virus
H9M6P4_9MONO RNA-directed RNA polymerase L, Ghana virus
A0A185KRV3_9MONO RNA-directed RNA polymerase L, Cedar virus
Q997F0| NIPAV RNA-directed RNA polymerase L, Nipah virus
O89344|L_HENDH RNA-directed RNA polymerase L, Hendra virus (isolate Horse/Australia/Hendra/1994)
G5CPP9_9MONO RNA-directed RNA polymerase L, *Henipavirus hendraense*
F4YH99_9MONO RNA-directed RNA polymerase L, *Henipavirus hendraense*

Figure 5: MSA of the NTD and RdRp domains of NiV and its related viruses

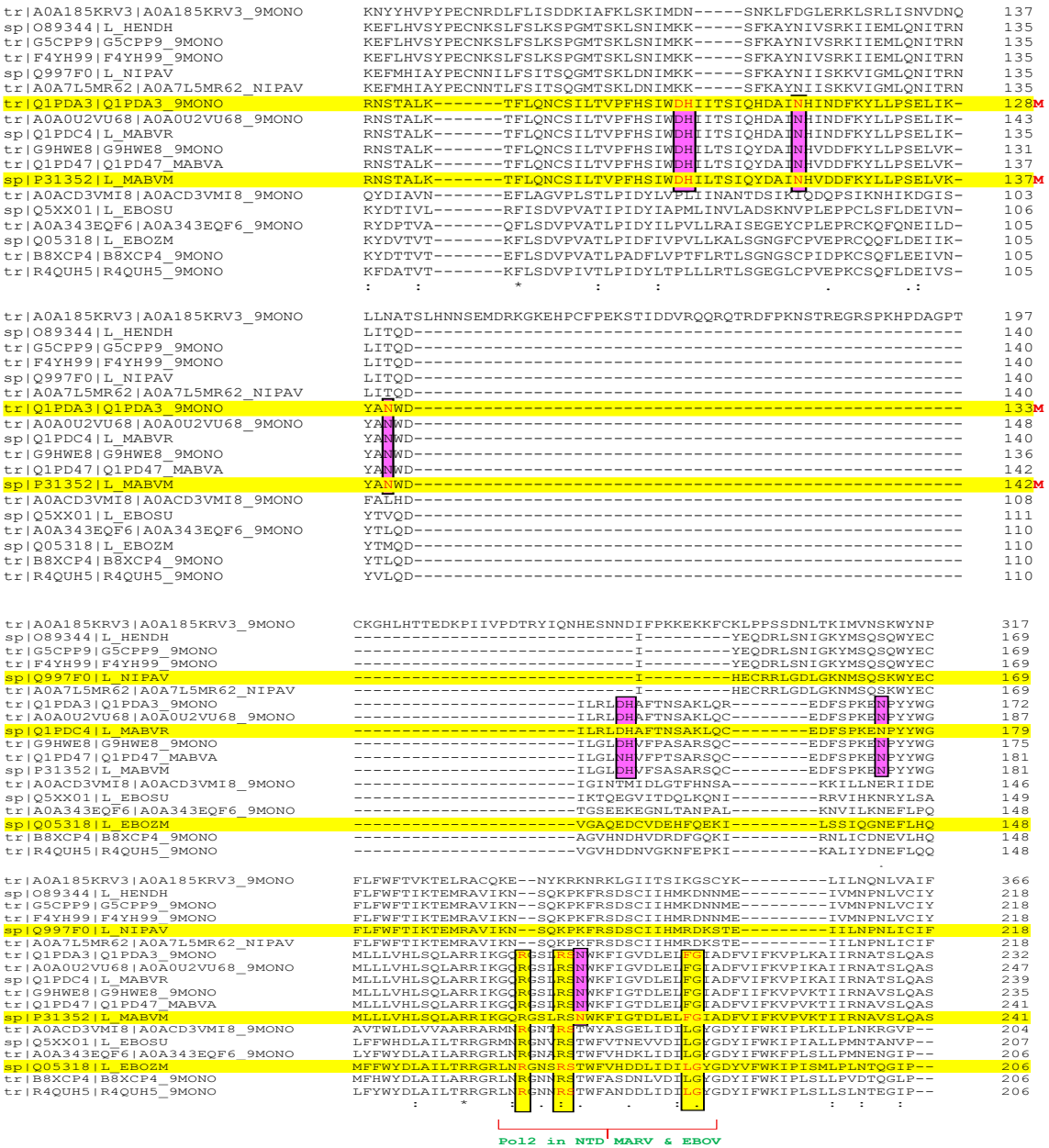
3.1 'Mix and Match' MSA analysis of the NTD and RdRp domains of all three bat-borne viruses: MARV, EBOV and NiV

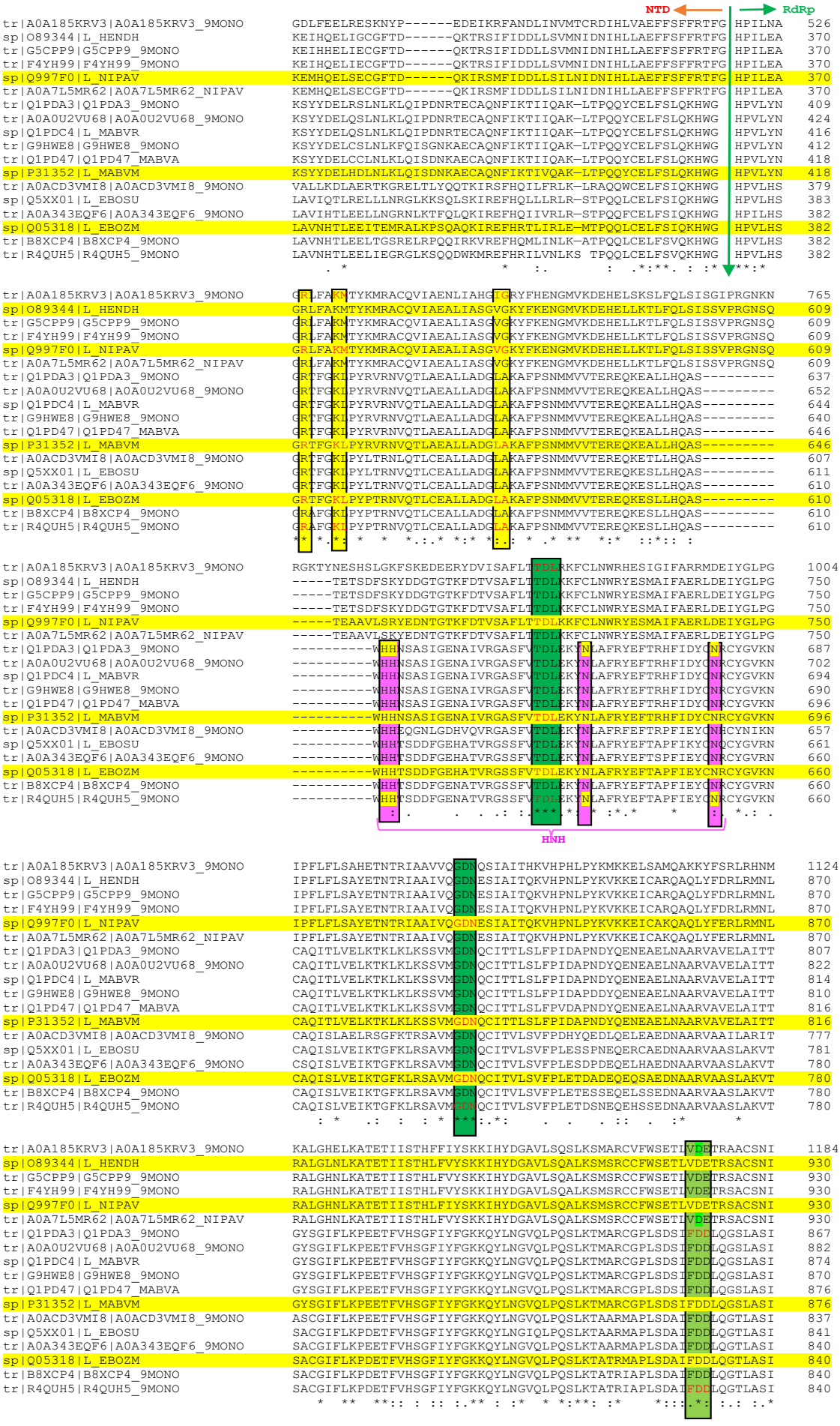
Figure 6 shows the 'Mix and Match' MSA of part of the NTD and RdRp domains of all the three bat-borne human pathogenic viruses, viz. MARV, EBOV and NiV (only the region required for the discussion is shown here). The representative strains of the three viruses are highlighted in yellow. The tentative RdRp domain in all three viruses is marked by arrows. The conserved motifs and amino acids in the catalytic core are highlighted in yellow. The typical, proton abstractor K and the nucleotide discriminator R at -4 from the catalytic K are completely conserved in all and the template-binding pair is highly conserved and is completely aligned from all three viral RdRps (Fig. 6). Only the MARV and EBOV harbour an additional putative polymerase catalytic core at the NTD (highlighted in yellow), but absent

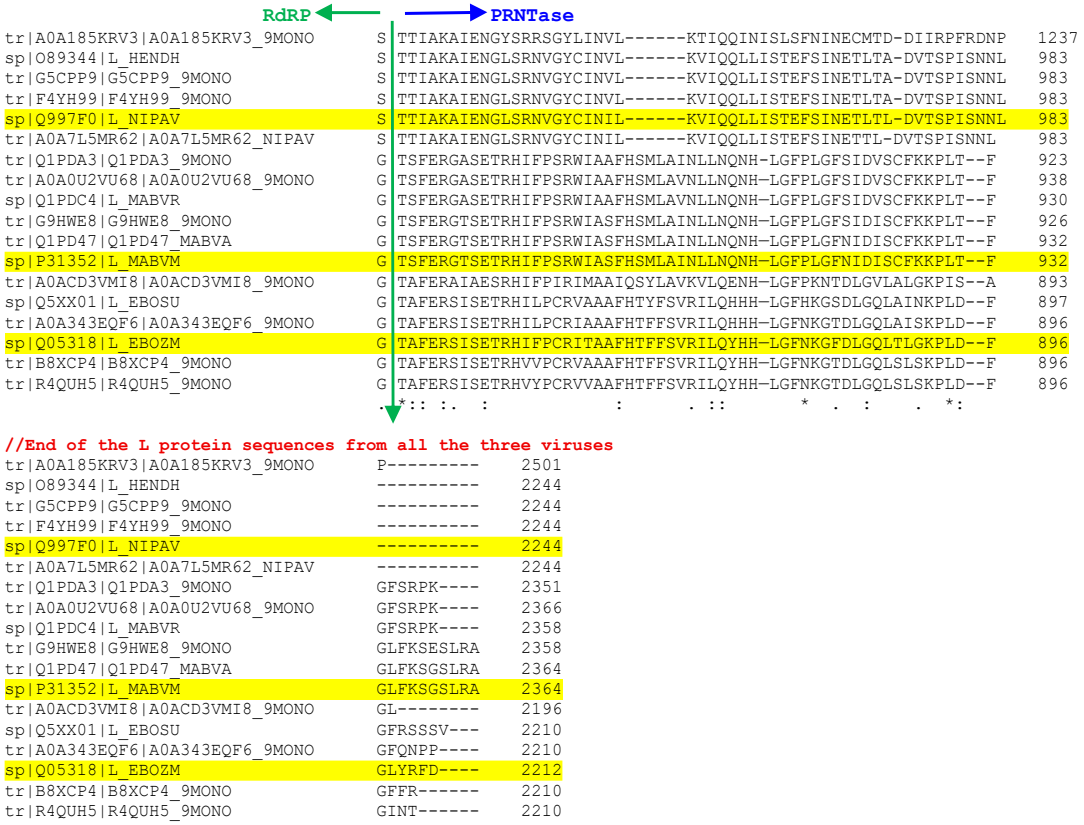
in the NiV. Another significant finding is the presence of an additional -DH- type putative HNH-endonuclease motifs (highlighted in magenta) in the NTD of MARV (highlighted in magenta), but not in the EBOV and NiV [22].

The HNH-endonucleases are essentially involved in a metal ion-mediated cleavage of DNAs resulting in single-stranded breaks (nicking) or double-stranded cuts. The HNH-endonuclease motifs are ubiquitous in biological systems and have been reported from both prokaryotes and eukaryotes. They are known to be involved in many important functions like, DNA packaging in double-stranded DNA viruses and herpesviruses, DNA homing where it acts as mobile genetic elements encoded by introns and inteins (where it creates double-stranded breaks on DNAs that drive the self-propagation and horizontal transfer of these elements), restriction, repair, chromosome degradation, etc. One of the main functions of these endonucleases is to promote the lateral transfer of their own coding and flanking DNA regions between genomes, by a recombination-dependent process known as 'homing' [22]. The presence of HNH-endonuclease motif has been already reported in the DNA polymerases of poxviruses like smallpox and monkeypox viruses, which are linear, double-stranded DNA viruses [30] and in the PB1 subunit of the RNA polymerase of human influenza viruses A, B and C which are *segmented*, negative-strand RNA viruses, in the CRISPR-Cas9 system (a bacterial defense mechanism against any invading phages or genetic elements), colicins and pyosins (types of bacteriocins), group I and group II introns, inteins, etc. by Palanivelu [22, 31].

CLUSTAL O (1.2.4) 'Mix and Match' MSA of the NTD and RdRp domains from MARV, EBOV and NiV and their related viruses.




NTD
RdRp
HNH



A0A185KRV3_9MONO RdRp of L protein, Cedar virus
O89344|L_HENDH RdRp of L protein, Hendra virus (isolate, Horse/Autralia/Hendra/1994)
G5CPP9_9MONO RdRp of L protein, *Henipavirus hendraense*
F4YH99|F4YH99_9MONO RNA-directed RNA polymerase L, *Henipavirus hendraense*
Q997F0|L_NIPAV RdRp of L protein, **Nipah virus**
A0A7L5MR62_NIPAV RdRp of L protein, Nipah virus

Q1PDA3_9MONO RdRp of L protein, Lake Victoria **Marburg virus**, (DRC1999)
A0A0U2VU68_9MONO RdRp of L protein, *Orthomarburgvirus marburgense*
Q1PDC4|L_MABVR RdRp of L protein, Lake Victoria Marburgvirus (strain Ravn-87)
G9HWE8_9MONO RdRp of L protein, Lake Victoria Marburgvirus (Leiden)
Q1PD47_MABVA RdRp of L protein, Lake Victoria Marburgvirus (strain Angola/2005)
P31352|L_MABVM RdRp of L protein, Lake Victoria Marburgvirus (strain Musoke-80)

A0ACD3VMI8_9MONO PdRP of L protein, *Cuevavirus lloviuense*
Q5XX01|L_EBOSU RdRp of L protein, Sudan Ebolavirus (strain Human/Uganda/Gulu/2000)
A0A343EQF6_9MONO RdRp of L protein, Bombali virus
Q05318|L_EBOZM RdRp of L protein, Zaire **Ebola virus** (strain Mayinga-76)
B8XCP4_9MONO RdRp of L protein, Tai Forest ebolavirus
R4QUH5_9MONO RdRp of L protein, Bundibugyo virus.

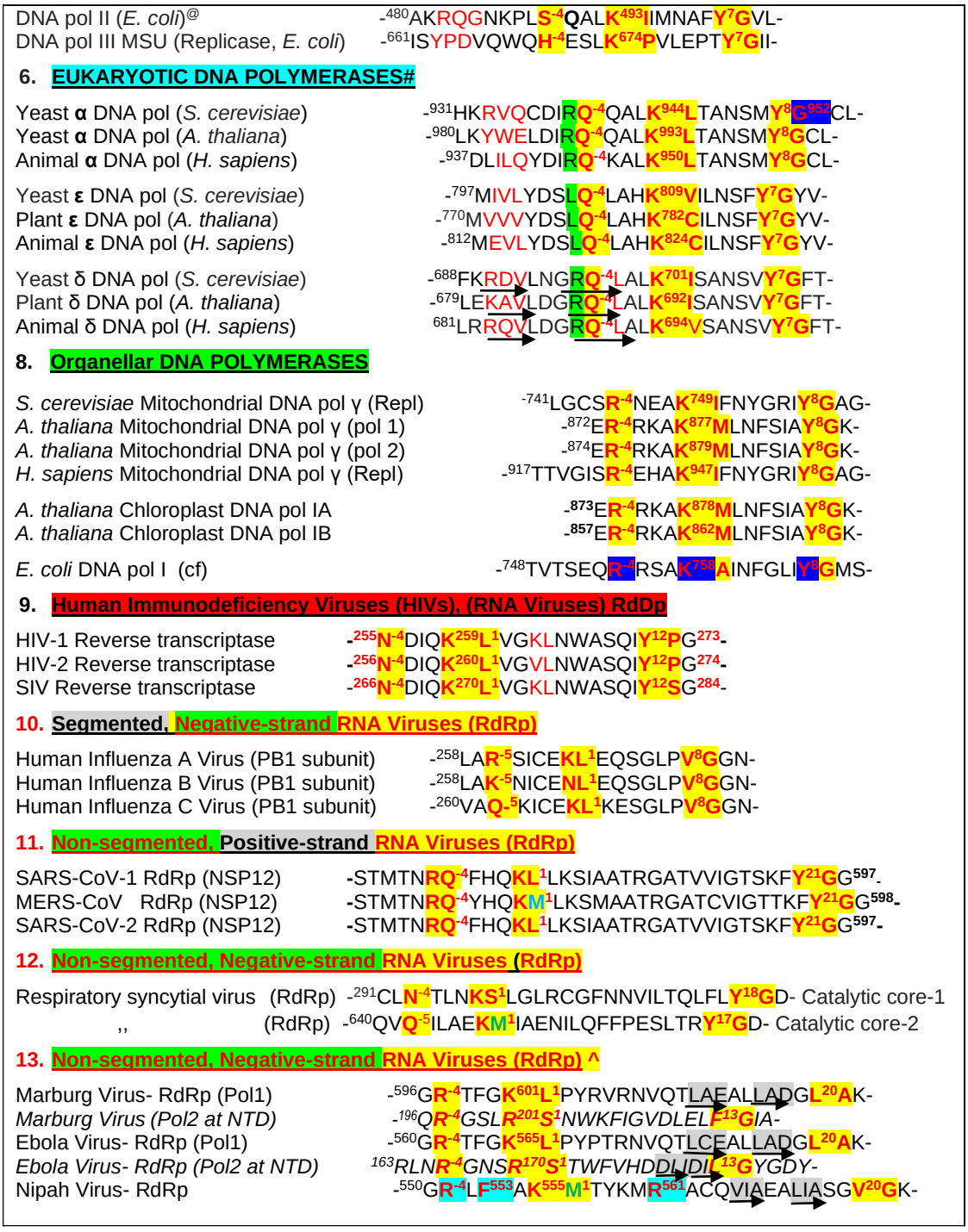
Figure 6: 'Mix and Match' MSA of Marburg, Ebola, Nipah and their related viruses

3.2 Conservation of the catalytic core region in the RNA/DNA polymerases from viruses, prokaryotic and eukaryotic organisms

Table 1 shows that the polymerase catalytic core is highly conserved in all organisms from viral (bacterial and animals), prokaryotic, (Eubacterial and Archaebacterial) and eukaryotic (plants and animals) RNA/DNA polymerases (only the representative organisms are shown in the Table 1). It is clear from the table, that invariably, all these polymerases possess three highly conserved catalytic core components, viz. a basic catalytic amino acid, R/K for proton abstraction from the growing 3'-OH to enable nucleotide addition, an R/Q/N at -4 from the proton abstracting K/R for nucleotide discrimination and a template-binding pair -YG- (the template-binding pair shows small variations). Interestingly, by sequence similarity, all three conserved elements are observed among the three human pathogenic bat-borne viruses, viz. MARV, EBOV and NiV.

Table 1: Conservation of polymerase catalytic region(s) in the viral, prokaryotic and eukaryotic RNA/DNA polymerases (Proposed and Confirmed).

Polymerase Type	Catalytic Core Region(s)
1. BACTERIAL VIRUSES	
SSU RNA POLYMERASES: SSU DNA-dep RNAPs	
Viral T7 SSU RNA pol	-^{620}WLAY^{-8}GVTR^{-4}SVTK^{631}RSVMTLAY^{8}GS-]
Viral SP6 SSU RNA pol	-^{612}WDSI^{-8}GITR^{-4}SLTKKPVMTLPY^{8}GS-
Viral T3 SSU RNA pol	-WLAY^{-8}GVTR^{-4}SVTKRSVMTLAY^{8}GS^{642}-
Viral BPK11 SSU RNA pol	-WLQY^{-8}GVTR^{-4}KVTKRSVMTLAY^{8}GS^{634}-
2. HUMAN VIRUSES	
(+) Strand RNA Viruses (SARS-Coronaviruses)	
RNA-dep RNAPs (NSP12)	
SARS-CoV-1	-STMTNR^{-5}Q^{-4}FHQKLLKSIAATRGATVVIGTSKFY^{21}GG^{597}-
SARS-CoV-2	-STMTNR^{-5}Q^{-4}FHQKLLKSIAATRGATVVIGTSKFY^{21}GG^{597}-
MERS-CoV	-STMTNR^{-5}Q^{-4}YHQKMLKSMAATRGATCVIGTTKFY^{21}GG^{598}-
3. PROKARYOTIC AND EUKARYOTIC MSU DNA-dep RNAPS (mRNAs)	
a) MSU RNAP Family (Prokaryotic and Eukaryotic Initiation subunits: (β and Rpb2 subunits)	
E. coli MSU RNAP β subunit	-^{539}TRER^{-6}AGFEVRDVHPTHY^{7}GRV^{558}-
S. cerevisiae MSU RNAP II Rpb2 subunit	-^{851}FR^{-5}SLFFRSYMDQEKKY^{9}GMS^{869}-
S. pombe MSU RNAP II Rpb2 subunit	-^{840}FR^{-5}SIFYRTYTDQEKKI^{9}GMT^{858}-
A. thaliana MSU RNAP II Rpb2 subunit	-^{819}FRSYR^{-5}DEEKKMGTLVKEDF^{9}GRP^{840}-
Human MSU RNAP II Rpb2 subunit	-^{806}FR^{-5}SVFYRSYKEQESKK^{9}GFDQ^{825}-
b) MSU RNAP Family (Prokaryotic and Eukaryotic Elongation subunits: (β' and Rpb1 subunits)	
E. coli MSU RNAP β' subunit	-^{833}NSV^{-6}DAVKVRSVVSC^{6}DTDFGVC^{13}AHC^{16}Y^{16}GRDL^{861}-
S. cerevisiae MSU RNAPII Rpb1 subunit	-^{55}DPR^{-6}LGSIDRNLKC^{5}QTC^{8}QEGMNEC^{15}PGHF^{18}GHI^{84}-
P. pastoris MSU RNAPII Rpb1 subunit	-^{55}DPK^{-6}LGSIDRNFKC^{5}QTC^{8}GEGMAEC^{15}PGHF^{18}GHM^{84}-
A. thaliana MSU RNAPII Rpb1 subunit	-^{54}DTR^{-7}LGTIDRKVKC^{4}ETC^{7}MANMAEC^{14}PGHF^{17}GYL^{83}-
H. sapiens MSU RNAPII Rpb1 subunit	-^{59}DPR^{-6}QGVIERTGRC^{5}QTC^{8}AGNMTEC^{15}PGHF^{18}GHI^{88}-
4. ORGANELLAR SSU RNAPs (NEPs)	
Mitochondrial (*S. cerevisiae*)	-TR^{-4}KVVKQTVMTNVY^{8}GV^{1024}-
Mitochondrial *(Arabidopsis thaliana)*	-QVDR^{-4}KLVKQTVMTSVY^{8}GV^{762}-
Mitochondrial *(Nicotiana sylvestris)*	-QVDR^{-4}KLVKQTVMTSVY^{8}GV^{788}-
Mitochondrial RNA pol (*H. sapiens*)	-TR^{-4}KVVKQTVMTVVY^{8}GV^{1001}-
Chloroplast *(Oryza sativa*, Japonica)	-QVDR^{-4}KLVKQTVMTSVY^{8}GV^{739}-
Chloroplast *(Arabidopsis thaliana)*	-QVDR^{-4}KLVKQTVMTSVY^{8}GV^{779}-
5. HUMAN and ANIMAL DNA VIRUSES (DNA-dependent RNAPs)	
Smallpox RNA pol (DdRp)	-SR^{-4}YPDRDSMVCHRILTY^{11}GK^{364}-
Mpox virus RNA pol (DdRp)	-SR^{-4}YPDRDSMVCHRILTY^{11}GK^{364}-
Cowpox virus RNA pol (DdRp)	-SR^{-4}YPDRDSMVCHRILTY^{11}GK^{364}-
6. VIRAL and PROKARYOTIC DNA POLYMERASES	
T7 phage DNA pol	-^{509}NQIAAELPTR^{-4}DNAK^{522}TFIYGFLY^{8}GAG
Smallpox DNA pol	-^{647}TEKAIYDSMQ^{-4}YTYK^{660}IANSVY^{7}GLM-
Vaccinia DNA pol	-^{648}TEKAIYDSMQ^{-4}YTYK^{661}IVANSVY^{7}GLM-
Mpox DNA pol	-^{648}TEKAIYDSMQ^{-4}YTYK^{661}IVANSVY^{7}GLM-
DNA pol I (*E. coli*)@	- ^{745}PLETVTSEQR^{-4}RSAK^{758}AINE^{762}GLIY^{8}GMS-
Taq pol I (*T. aquaticus*)	-^{650}PREAVDPLMR^{-4}RAAK^{663}TINFGVLY^{8}GM-
Pfu DNA pol (*P. furiosus*)	-^{478}ILLDYRQ^{-4}KAIK^{488}LLANSFY^{7}GYYGYAK-
Desulfurococcus DNA pol	-^{481}YRQ^{-4}RAIK^{487}ILANSYY^{7}GYYAYA-
Thermococcus gorgonarius DNA pol	-^{481}YRQ^{-4}RAIK^{487}ILANSFY^{7}GYYGYA-



Palanivelu and references therein [32].

- Pol, Polymerase; NEPs, Nuclear Encoded Polymerases; Repl, Replicase.
- *Pfu, Pyrococcus furiosus;* SIV, Simian Immunodeficiency Virus
- The confirmed active site amino acids (by SDM, X-ray crystallography and active site-directed inhibitors) are highlighted in dark blue and those from the cryo-EM data are highlighted in light blue.
- #Eukaryotic replicative DNA polymerases, viz. α pols act as primase, whereas the ε and δ pols involve in leading and lagging strand syntheses, respectively.
- @ DNA pol I and II are not the replicative pols, but only pol II possesses the two typical conserved motifs, -SLYPS- and –YGDTDS- as those in eukaryotic replicative DNA pols α and δ suggesting a separate evolutionary pathway. It should be noted that these two motifs are also found in the poxviral DNA pols. However, they are not found in the bacterial DNA pol I, bacterial replicative pol III and in the eukaryotic ε pol, suggesting their separate evolutionary pathway. In *E. coli* DNA pol I, replacement of the conserved catalytic core amino acid, K^{758} with A^{758} caused a 1,000-fold reduction in k_{cat} In *Taq* pol I, the conserved catalytic core amino acids, R^{659} and K^{663} were immutable.
- ^Present work. A tripeptide tandem repeat (marked by arrows) is seen within the proposed polymerase catalytic core regions of MARV, EBOV and NiV.

3.3 Conservation of the catalytic metal-binding sites in the RNA/DNA polymerases from viruses, prokaryotic and eukaryotic organisms

Table 2 shows the conservation of the catalytic metal-binding sites of the RNA/DNA polymerases from viruses to humans (only the representative organisms are shown here). In all these polymerases, the metal-binding sites are invariably composed of two to three invariant Ds. The *positive-strand* RNA viruses use -S/ADD- and -GDD- motifs for catalytic Mg^{2+}-binding, whereas the *negative-strand* RNA viruses use -SDD- and -GDN- motifs. The three bat-borne viruses use -GDN- and -TDL- motifs (arrived at by sequence similarity and in the case of EBOV the metal-binding site amino acids are also confirmed by SDM experiments). HIVs use -LDV- and -MDD- in the catalytic metal-binding site. Interestingly, the DdDps from poxviruses and eukaryotic replicative polymerases (α and δ) use the same metal-binding sites having invariant Ds. The SSU RNA polymerases from mitochondria, chloroplasts and bacteriophages use -DGS/T- and -HDS- as metal-binding sites. Based on the highly conserved catalytic Mg^{2+}-binding sites, the DdDps (EC 2.7.7.7) maybe broadly classified into two evolutionary families: one with the -YGDTD- type with an invariant -SLYPS- motif and the other with the -DTD- type. Both types use one more invariant D, as shown in Table 2. DNA polymerases from bacteriophages, bacterial pol II type, poxviruses, eukaryotic replicative α and δ pols. pox viral DNA pols, etc. belong to the -YGDTD- type, whereas the eubacterial pol I type, archaebacterial and eukaryotic replicative ε pol, etc. belong to the -DTD- type, further corroborating their distinct evolutionary pathway. It is interesting to note that both types do not differ much in their catalytic core amino acids. The single-subunit (SSU) organellar and bacteriophage RNA polymerases use different types of metal-binding sites, viz. -Q/FDGS/T- and -HDS- motifs.

Table 2: Conserved catalytic metal-binding sites in various RNA/DNA polymerases (RdRps, RdDps DdDps and DdRps)

Organism	Metal-binding sites	
Positive-strand RNA Viruses, RdRps (EC 2.7.7.48)		
SARS-CoV-1	-LSD^{760}DA- &	-QGDD^{825}Y-
MERS-CoV	-LSD^{761}DG- &	-DGDD^{826}G-
SARS-CoV-2	-LSDD^{761}A- &	-QGDD^{826}Y-
Bat_RaTG13-CoV	-LSDD^{761}A- &	-QGDD^{825}Y-
Pangolin CoV	-LSDD^{761}A- &	-QGDD^{826}Y-
Polio Type 1 (Subtn)	-IDLD^{523}H- &	-YGDD^{533}V-
Hepatitis C (Genotype 1B, Japanese)	-DSDD^{275}G- &	-NGDD^{2738}L-
Dengue	-YADD^{633}TA- &	-SGDD^{664}CV-
Zika	-YADD^{635}TA- &	-SGDD^{666}CV-
Yellow Fever	-YADD^{284}TA- &	-SGDD^{117}CV-
Rubella	-LGDD^{28}A- &	-QGDD^{887}M-
Encephalomyocarditis	-YDVD^{2065}Y- &	-YGDD^{2163}L-
Negative-strand RNA Viruses, RdRps (EC 2.7.7.48)		
Influenza A	-TGD^{305}NTK- &	-QSSDD^{446}FA-
Influenza B	-TGD^{305}NTK- &	-QSSDD^{445}FA-
Influenza C	-TGD^{307}NSK- &	-QSSDD^{451}FV-
Rabies	-QGD^{729}NQV- &	-LESDD^{1589}NI-
Measles	-QGD^{773}NQT- &	-AYGDD^{1256}DS-
Mumps	-QGD^{779}NQA- &	-SSGDD^{1195}KF-
Negative-strand RNA Viruses, RdRps (EC 2.7.7.48)^		
Marburg (Musoke-80)	-MGD^{740}NQC- &	-VTD^{936}LE-
Ebola (Zaire)	-MGD^{742}NQC- &	-VTD^{832}LE-
Nipah	-QGD^{832}NES- &	-TTD^{722}LK-
Human Immunodeficiency Viruses (HIVs) RdDps, (EC 2.7.7.49)		
HIV-1 Reverse transcriptase	-LD^{110}V-	-YMDD^{185}I-
HIV-2 Reverse transcriptase	-LD^{110}V-	-YMDD^{186}L-
SIV Reverse transcriptase	-LD^{121}I-	-YMDD^{197}L-
Poxviruses (DdDps, EC 2.7.7.7)		
Smallpox Virus (Vaccinia virus)	-IFD^{548}YN-	-VYGDTD^{753}SV-
Variola Virus	-IFD^{548}YN-	-VYGDTD^{752}SV-
Monkeypox Virus	-IFD^{549}YN-	-VYGDTD^{753}SV-
Other Viral, Prokaryotic and Eukaryotic Organisms (DdDps, EC 2.7.7.7)		
Bacteriophage (T4)	-^{408}SFD*LT-	-^{615}AAGDTDSV-
Bacterial DNA pol II (*E. coli*)	-^{417}VLDYK-	-^{542}IYGDTDST-
Yeast replicative α DNA Pols (*Sc*)	-^{882}VMDFN-	-^{993}VYGDTDSVM-
Plant replicative α DNA Pols (*Os*)	-^{924}LLDFN-	-^{1052}IYGDTDSIM-
Animal replicative α DNA Pols (*Hs*)	-^{858}LLDFN-	^{999}IYGDTDSIM-
Yeast replicative δ DNA Pols (*Sc*)	-^{808}TLDFN-	-^{759}VYGDTDSV-
Plant replicative δ DNA Pols (*At*)	^{599}TLDFA-	-^{750}IYGDTDSV-
Animal replicative δ DNA Pols (*Hs*)	-^{600}TLDFS-	-^{752}VYGDTDSV-
*The first invariant D is found just before the conserved -SLYPS- motif.		
Eubacteria (EC 2.7.7.7)		
Bacterial DNA pol I (*E. coli*)	-YD^{424}R,	ED^{501}AD, -FDTE^{357}T-
Bacterial DNA pol III, Repl (*E. coli*)	-FD^{555}F-	-PDFD^{402}V-
***Eukaryotes* (EC 2.7.7.7)**		
Eucaryotic DNA pol ε, Repl (Yeast, *Sc*)	-YHVD^{640}V-	—LDTD^{877}G
Eucaryotic DNA pol ε, Repl (Human, Hs)	-YHLD^{626}V-	—LDTD^{862}G-
Eucaryotic DNA pol ε, Repl (Plants, *At*)	-YHLD^{598}V-	—LDTD^{836}G-
Archaebacteria (EC 2.7.7.7)		
Thermococcus gorgonarius DNA pol	-VYLD^{404}F-	-ADTD^{542}G-
Pfu DNA pol (*P. furiosus*)	-VYLD^{405}F-	-IDTD^{543}G-
Desulfurococcus DNA pol	-IYLD^{404}Y-	-ADTD^{542}G-

Organism	Metal-binding sites
Mitochondria (SSU RNA pol, EC 2.7.7.6)	
Arabidopsis thaliana	-QD^{677}GSC- & -VH^{908}DSF-
Nicotiana sylvestris	-QD^{703}GSC- & -VH^{934}DSY-
S. cerevisiae	-QD^{945}GTC- & -VH^{1188}DSY-
Candida albicans	-QD^{881}GTC- & -VH^{1121}DSF-
Neurospora crassa	-QD^{901}GTC- & -VH^{1180}DSF-
Chloroplast (SSU RNA pol, EC 2.7.7.6)	
Arabidopsis thaliana	-QD^{694}GSC- & -VH^{925}DSY-
Brassica napus	-QD^{680}GSC- & -VH^{912}DSY-
Bacteriophages (SSU RNA pol, EC 2.7.7.6)	
T7 phage	-FD^{537}GSC- & -HD^{812}SF-
Lambda Phage	-FD^{537}GSC- & -HD^{812}SF-

Palanivelu [5], and references therein.; Pol, polymerase; Repl, Replicase. ; *Sc, Saccharomyces cerevisiae; Os, Oryza sativa; Hs, Homo sapiens; At, Arabidopsis thaliana; P. furiosus, Pyrococcus furiosus;* Most of the metal-binding sites were confirmed by X-ray crystallography, cryo-EM, SDM and other techniques. ; ^ Present work. The metal-binding sites confirmed by SDM are highlighted in dark blue and by cryo-EM in light blue.

4 CONCLUSION

While a number of experimental therapeutics against these highly pathogenic human viruses target the viral RdRp, there are still significant gaps in our knowledge on this crucial viral enzyme, which is indispensable for the virus lifecycle as it mediates the two crucial activities, viz. replication and transcription. Until now, the experimental evidences are not complete in identifying the catalytic centre(s) of these viral RdRps. The present findings reveal for the first time that these highly pathogenic, human bat-borne RNA viruses, viz. MARV, EBOV and NiV, though they belong to two different families, yet possess the same structural organization of their RdRp and exhibit similar function(s) in performing the two essential activities, viz. transcription and replication. The high degree of conservation observed within the active-site architecture provides strong support for the development of polymerase-targeted antiviral agents capable of inhibiting viral multiplication and limiting the spread of these deadly pathogens in humans. Furthermore, the highly conserved active site structures of the RdRps of these viruses highlight their evolutionary significance and suggest that these catalytic domains have been highly conserved throughout evolution. Such comparative structural and functional analyses of RdRps of human pathogenic RNA viruses can potentially deliver a common platform for the rational design of broad-spectrum antiviral therapeutics. Future experimental studies, including SDM and other techniques of the predicted polymerase catalytic residues are warranted to validate their functional roles during polymerization. The apparent absence of the putative HNH-endonuclease motif in the NiV RdRp is intriguing.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments

The author wishes to thank Dr. N. Srinivasan, Former Professor, Department of Endocrinology, Post Graduate Institute of Basic Medical Sciences, University of Madras, Chennai, for useful suggestions on the manuscript.

Disclosure of conflict of interest

The author declares no conflict of interest.

REFERENCES

[1] Poltronieria P, Sunb B, Mallardoc M. RNA Viruses: RNA roles in pathogenesis, coreplication and viral load. Curr Genomics. 2015; 16: 327-335.

[2] Mohanty P, Panda P, Acharya RK, Pande B, Bhaskar L, Verma HK. Emerging perspectives on RNA virus-mediated infections: From pathogenesis to therapeutic interventions. World J Virol. 2023; 12: 242-255.

[3] Palanivelu P. DNA polymerases – An insight into their active sites and mechanism of action, In: Recent Advances in Biological Research, Vol 1, pp 1-39, SCIENCEDOMAIN International Book Publishers, UK. ISBN: 9788193422441, DOI: 10.9734/bpi/rabr/v1; 2019.

[4] Palanivelu P. Single Subunit RNA Polymerases: An Insight into their Active Sites and Catalytic Mechanism, In: Advances and Trends in Biotechnology and Genetics. Vol 1, pp 1-38, SCIENCEDOMAIN International Book Publishers, UK. Print ISBN: 978-93-89246-59-9, DOI:10.9734/bpi/atbg/v1; 2019.

[5] Palanivelu P. An overview of the proofreading functions in bacteria and SARS-Coronaviruses. Int J Biochem Res Rev. 2021; 30: 33-62.

[6] Palanivelu P. An Insight into the Active sites of the RNA Polymerase and Proofreading Exonuclease of the Human Respiratory Syncytial Virus. World J Adv Res Rev. 2023; 18: 842–858

[7] Palanivelu P. Identification of a Proofreading Domain within the Guanosine Diphosphate Polyribonucleotidyltransferase Domain in the Bat-borne Deadly Human Viruses: Marburg, Ebola and Nipah. World J Adv Res Rev. 2026; 30: 2246-2261.

[8] Cross RW, Mire CE, Feldmann H, Geisbert TW. Postexposure treatments for Ebola and Marburg virus infections. Nat. Rev. Drug Discov. 2018; 17: 413–434.

[9] Muhlberger E. Filovirus replication and transcription. Future Virol. 2007; 2: 205–215.

[10] Hume AJ, Muhlberger E. Distinct Genome Replication and Transcription Strategies within the Growing Filovirus Family. J Mol Biol. 2019; 431: 4290–4320.

[11] Prins KC, Binning JM, Shabman RS, Leung DW, Amarasinghe GK, Basler CF. Basic Residues within the Ebolavirus VP35 Protein Are Required for Its Viral Polymerase Cofactor Function. J. Virol. 2010; 84: 10581–10591.

[12] Muhlberger E, Weik M, Volchkov VE, Klenk HD, Becker S. Comparison of the transcription and replication strategies of Marburg virus and Ebola virus by using artificial replication systems. J Virol. 1999; 73: 2333-2342.

[13] Peng Q, Dong Y, Jia M, Liu Q, Bi Y, Qi J, Shi Y. Cryo-EM structure of Nipah virus L-P polymerase complex. Nat. Commun. 2024; 15:10524. doi.org/10.1038/s41467-024-54994-5

[14] Uchida S, Hor R, Sato H, Kai C, Yoneda M. Possible Role of the Nipah Virus V Protein in the Regulation of the Interferon Beta Induction by Interacting with UBX Domain-Containing Protein1. Nature Sci Rep. 2018; 8:7682. doi: 10.1038/s41598-018-25815-9.

[15] Harcourt BH, Tamin A, Ksiazek TG, Rollin PE, Anderson LJ, Bellini WJ, Rota PA. Molecular characterization of Nipah virus, a newly emergent paramyxovirus. Virology. 2000; 271: 334–349.

[16] Li G, Du T, Wang J, Jie K, Ren Z, Zhang X, Zhang L, Wu S, Ru H. Structural insights into the RNA-dependent RNA polymerase complexes from highly pathogenic Marburg and Ebola viruses. Nat Commun. 2025; 16:3080. https://doi.org/10.1038/s41467-025-58308-1.

[17] Yuan B, Peng Q, Cheng J, Wang M, Zhong J, Qi J, Gao GF, Shi Y. Structure of the Ebola virus polymerase complex. Nature. 2022; 610:394-401.

[18] Peng Q, Yuan B, Cheng J, Wang M, Gao S, Bai S, Zhao X, Qi J, Gao GF, Shi Y. Molecular mechanism of *de novo* replication by the Ebola virus polymerase. Nature. 2023; 622:603-610. doi: 10.1038/s41586-023-06608-1.

[19] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soeding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011; 7, 539: 1-6.

[20] Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 2017; doi.org/10.1002/pro.3290.

[21] Jacob ST. Crozier I, Fischer WA, Hewlett A, Kraft CS, de La Vega M, Soka MJ, *et al*. Ebola virus disease. Nat. Rev. Dis. Primers. 2020; 6, 13.

[22] Palanivelu P. Assessment and Analyses of Homing Endonucleases and Mechanism of Action of CRISPR-Cas9 HNH Endonucleases, In Current Advances in Chemistry and Biochemistry Vol. 1. Chapter-3, pp 20-48. SCIENCEDOMAIN International Book Publishers, UK. Print ISBN: 978-93-90516-10-0, eBook ISBN: 978-93-90516-18-6. DOI: 10.9734/bpi/cacb/v1/2273E, 2021.

[23] Schmidt ML, Hoenen T. Characterization of the catalytic center of the Ebola virus L polymerase. PLoS Negl Trop Dis. 2017; 11:1-14.

[24] Tchesnokov EP, Raeisimakiani P, Ngure M, Marchant D, Götte M. Recombinant RNA-Dependent RNA Polymerase Complex of Ebola Virus. Nature Sci Rep. 2018; 8:3970-3979.

[25] Wang Y, Zhao L, Zhang Y, Wang Y, Tang J, Liu S, Gao H, *et al.* Cryo-EM structure of Nipah virus RNA polymerase complex. Sci. Adv. 2024; 10, eadr7116.

[26] Peng, Q., Dong, Y., Jia, M. Liu Q, Bi Y, Qi J, Shi. Y Cryo-EM structure of Nipah virus L-P polymerase complex. Nat Commun. 2024; 15: 10524. doi.org/10.1038/s41467-024-54994-5.

[27] Sala FA, Hillen HS. Structure and dynamics of the Nipah virus RNA-dependent RNA polymerase. Trends in Biochemical Sciences. 2026: xx: 1-13 DOI: 10.1016/j.tibs.2026.06.008.

[28] Hu S, Kim H, Yang P, Yu Z, Ludeke B, Mobilia S, Pan J, Stratton M, Bian Y, Fearns R, Abraham J. Structural and functional analysis of the Nipah virus polymerase complex. Cell. 2025; 188: 688–703.

[29] Balıkçı E, Günl F, Carrique L, Keown JR, Fodor E, Grimes JM. Structure of the Nipah virus polymerase complex. The EMBO Journal. 2025; 44:563-586. doi: 10.1038/s44318-024-00321-z.

[30] Palanivelu P. Insight into the Replication Mechanism of Smallpox and Monkeypox Viruses. World J Adv Res Rev. 2024; 23: 2239-2260.

[31] Palanivelu P. An Insight into the Active Sites of the Catalytic Basic Protein Subunit PB1 of the RNA Polymerase of Human Influenza Viruses, World J Adv Res Rev. 2-023; 17: 625–656.

[32] Palanivelu P. Evolution and Adaptation of Enzymes' Active Sites: Enzymes Use Functionally Similar Active Site for Same Type of Reactions with Structurally Similar Substrates Irrespective of the Organism. World J Adv Res Rev. 2025; 28: 1810-1835.