International Journal of Biochemistry Research & Review

29(10): 29-52, 2020; Article no.IJBCRR.64238
ISSN: 2231-086X, NLM ID: 101654445

RNA Dependent RNA Polymerases of Severe Acute Respiratory Syndrome-Related Coronaviruses- An Insight into their Active Sites and Mechanism of Action

Peramachi Palanivelu[1*]

[1]Department of Molecular Microbiology, School of Biotechnology, Madurai Kamaraj University Madurai – 625 021, India.

Author's contribution

The sole author designed, analysed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/IJBCRR/2020/v29i1030236
Editor(s):
(1) Dr. Chunying Li , Georgia State University, USA.
(2) Dr. Varun Gupta, Albert Einstein College of Medicine, USA.
(3) Mohamed Fawzy Ramadan Hassanien, Zagazig University, Egypt.
Reviewers:
(1) Antonio Carlos Pereira de Menezes Filho, Goiano Federal Institute of Goiás, Brazil.
(2) Érica Tex Paulino Fundação Oswaldo Cruz (FIOCRUZ), Brazil.
(3) Karen Elizabeth Campos Correa, Universidad Peruana Cayetano Heredia (UPCH), Perú.
(4) Kartika Rathore, JNVU, India.
(5) Ndoe Guiaro Marcellin, Institute of Medical Research And Medicinal Plants Studies, Cameroon.
Complete Peer review History: http://www.sdiarticle4.com/review-history/64238

Original Research Article

Received 10 December 2020
Accepted 30 December 2020
Published 31 December 2020

ABSTRACT

Aim: To analyze the RNA-dependent RNA polymerases (RdRps) of Severe Acute Respiratory Syndrome (SARS)-related coronaviruses (CoVs) to find out the conserved motifs, metal binding sites and catalytic amino acids and propose a plausible mechanism of action for these enzymes, using SARS-CoV-2 RdRp as a model enzyme.
Study Design: Bioinformatics, Biochemical, Site-directed mutagenesis (SDM), X-ray crystallographic and cryo-Electron microscopic (cryo-EM) data were analyzed.
Methodology: Bioinformatics, Biochemical, Site-directed mutagenesis, X-ray crystallographic and cryo-EM data of these enzymes from RNA viral pathogens were analyzed. The advanced version of Clustal Omega was used for protein sequence analysis of the RdRps.
Results: Multiple sequence alignment (MSA) of RdRps from different SARS-related CoVs show a large number of highly conserved motifs among them. Though the RdRp from the Middle Eastern

**Corresponding author: E-mail: ppmkupp@gmail.com;*

Respiratory syndrome (MERS)-CoV differed in many conserved regions yet the active site regions are completely conserved. Possible catalytic regions consist of an absolutely conserved amino acid K, as in single subunit (SSU) RNA polymerases and most of the DNA dependent DNA polymerases (DdDps). The invariant 'gatekeeper/DNA template binding' YG pair that was reported in all SSU DNA dependent RNA polymerases (DdRps), prokaryotic multi-subunit (MSU) DdRps and DdDps is also highly conserved in the RdRps of SARS-CoVs. The universal metal binding motif –GDD- and an additional motif–SDD- are also found in all SARS-CoV RdRps. In stark contrast, the (–) strand RNA viral pathogens like Ebola, rabies, etc. use –GD**N**- rather than –GD**D**- for catalytic metal binding. An invariant YA pair (instead of an YG pair) is found in the primases of the SARS-CoVs. The SARS-CoVs RdRps and primases exhibit very similar active site and catalytic regions with almost same distance conservations between the template binding YG/YA pair and the catalytic K. In SARS-CoV RdRps an invariant R is placed at -5 which is shown to play a role in nucleoside triphosphate (NTP) selection and is in close agreement with SSU DdRps (viral family) and DdDps. In primases no such invariant R/K/H is found very close to the catalytic K in the downstream region, as found in RdRp and Nidovirus RdRp-Associated Nucleotidyltransferase (NiRAN) domains. An invariant YA pair is placed in the NiRAN domain instead of an YG pair, and an invariant H is placed at -5 position. Moreover, the Zn binding motif with the completely conserved Cs and a few DxD/DxxD type metal binding motifs are found in the RdRps and NiRAN domain. However, the primases contained only the DXD type metal binding motifs.

Conclusions: The SARS and SARS-related CoV RdRps are very similar as large peptide regions are highly conserved among them. The closer identity between the RdRps of palm civet-CoV and SARS-CoV (CoV-1) suggest their possible link as found between their spike proteins also. The invariant YG and KL pairs may play a role in template binding and catalysis in SARS-CoV RdRps as reported in DdDps. An additional invariant –YAN- motif found in SARS-CoV RdRps may play a crucial role in nucleotide discrimination.

Keywords: Coronaviruses; RNA-dependent RNA polymerases; primases; NiRAN domain; polymerase conserved motifs; active sites; catalytic mechanism.

1. INTRODUCTION

Large numbers of human and animal pathogens belong to the positive-strand RNA viruses which have been causing major global health-care crises and unprecedented economic losses. The recent pandemic, caused by SAR-CoV-2, affected more than 72.0 million people resulting in ~1.6 million deaths worldwide. Both vaccines and antivirals have been developed at warp speed to contain the spread of the virus. Whereas the vaccines are targeted to the spike protein, the antivirals are mainly targeted to the crucial enzymes that participate in its lifecycle. One of main antivirals is against the crucial enzyme RdRp of the viruses. These RNA viruses employ the enzyme RdRp for multiplication in animal and human cells. The RdRp performs both transcription and replication processes in all these RNA viruses (except retro viruses). Therefore, RdRps have been the main target for antiviral drug development to control these RNA viruses including SARS-CoV-2. However, these efforts are hindered by limited structural information on the RdRp catalytic core and its catalytic mechanism. Though RNA viruses are highly divergent in nature, yet their key enzyme, viz. the RdRp is structurally and functionally found to be very similar. For example, all of them are structurally similar and exhibit a characteristic cupped right-hand with three major domains, viz. Fingers, Palm and Thumb, similar to DNA polymerases and SSU RNA polymerases. Furthermore, all of them perform the same function as nucleotidyltransferases indicating inextricable relationship between structure and function among them. These viral RdRps are mostly single subunit enzymes but are associated with or part of other viral enzymes related to viral replication process like primases, methyltransferases, proofreading exonucleases, etc. Unlike SARS-CoV RdRps, many viral RdRps do not possess the proofreading activity. It is interesting to note that RdRp activity has also been demonstrated in plants, fungi and nematodes where this activity was shown to be essential for gene silencing events (co-suppression in plants, quelling in fungi and RNA interference in nematodes) [1]. An in-depth analyses and characterization of the SARS-CoV-2 genome and its proteome will pave way for the development of anti-SARS-CoV strategies including effective and safe vaccines and effective antivirals to contain the present and future pandemics.

1.1 Dynamics of Viral RNA Synthesis

Viral RdRps (EC 2.7.7.48) belong to the Main class 'Transferases' and are involved in the transfer of ribonucleoside triphosphates (NTPs) to the growing 3'–OH end of the nascent RNA chain (Fig. 1). The NTPs are added complementary to the template strand, one at a time. The overall reaction catalyzed by an RdRp may be written as,

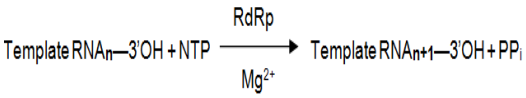

Fig. 1. RNA polymerization reaction by RNA polymerases

(A Mg^{2+} ion (invariably coordinated by t*wo dicarboxalic acids, viz. D-D or D-E) binds* specifically to the β- and γ-phosphates of incoming *NTPs and acts as "charge shielder"* and orients the α-phosphate for polymerization. *The hydrolysis of P—O bond results in release of large amount of negative free energy, which is efficiently utilized in the formation of a new phosphodiester bond*).

RdRps read the template RNA from 3'→5' direction and therefore, a new RNA strand is synthesized in the 5'→3' direction. RNA polymerases after successful initiation, elongate the 3'-end of the RNA by the addition of one nucleotide at a time. Unlike DNA polymerases, they can initiate a chain *'de novo'* without a primer. However, some viral RdRps use an RNA or protein primer for initiation.

When they use RNA primers, they are capped at their 5'-end like the cellular mRNAs either by the RdRp associated activity or snatched from the cellular mRNAs by a 'cap snatching mechanism.'

The RNA synthesis by RdRps involves three steps, viz. initiation, elongation and termination. The newly formed RNA copies serve as blueprints for further replication and transcription processes. In general, in SARS-related CoVs, the viral RNA is first translated into a polyprotein by the host machinery, which is subsequently cleaved by the proteases embedded in the polyprotein itself releasing the RdRp. In SARS-CoVs, the RdRp catalyzes the synthesis of the RNA genome by using the (+) RNA strand as a template to produce a complementary (−) RNA strand starting from 3′-end.

1.2 Viral RNA Synthesis: Initiation, Elongation and Termination

Correct initiation of RNA synthesis is crucial for the integrity of the viral genome. These viruses are found to use two different mechanisms for initiation of viral genome replication; i) *de novo* or primer-independent initiation and ii) primer-dependent initiation. In *de novo* primer-independent initiations (e.g., Flaviviruses and Noroviruses), one initiation nucleotide, usually a G at +1, provides the 3'-OH as the primer for the addition of the next nucleotide at +2, whereas the primer-dependent initiation is of two types: i) either use an oligonucleotide primer (usually synthesized by an RdRp associated primase (e.g., CoVs) or ii) a protein primer (e.g., Picornaviruses) where the 5′ extremity of the viral RNA is covalently linked to an ~20 amino acid Viral Protein genome-linked peptide called VPg. In this type of initiation, a conserved Tyr-OH of the VPg is projected into the enzyme active site and thus, the initiation starts with Tyr (the Tyr-OH essentially mimics the 3'-OH for the formation of phosphodiester bond with the first nucleotide for further elongation). Protein priming can also use an invariant Ser/Thr in the VPg linked peptides [2]. Capped RNA structures, captured from cellular mRNAs by 'cap snatching' also serve to initiate replication especially in Orthomyxo- and Bunya-viruses [3,4].

For transcription process, the viral polymerase binds to the genomic RNA at the 3' non-coding leader promoter and transcribes subsequently all viral mRNAs. After successful initiation, the elongation complex takes over and completes the replication/transcription cycle. Capping (guanine-N-7 methylation) is concomitant with initiation of mRNA transcription. Polyadenylation of mRNAs occurs by a stuttering mechanism at a slippery stop-site present at the end viral genome. By this mechanism, a U on the genomic template is copied hundreds of times at the end of viral mRNA producing a poly-(A) tail. Switching of the RdRp from transcription to the replication mode is mainly dependent on the intracellular concentrations of the nuclear capsid protein (during replication mode it ignores the transcription signals).

1.3 Error Rates and Proofreading in RdRps

Unlike the prokaryotic and eukaryotic RNA polymerases, the viral RdRps make high rate of errors (~10^{-4}) during copying the genome. This is

mostly due to lack of proofreading activity in most of these RNA viruses. The lack of proofreading activity also favours constant creation of new genetic variants to adapt to constantly changing environmental conditions. These mutations may result in loss-of-function or gain-of-function, where both are exploited by these viruses for their successful adaptation to new environments and hosts [5]. In contrast, the RdRps of CoVs maintain their genome integrity during replication by employing an associated proofreading activity. For viruses, whose survival mainly rely on the efficient production of many progeny copies, adaptability, stability and velocity of replication are essential and these features are slowly acquired during prolonged evolutionary times by mutation and selection. As mentioned earlier, the viral polymerases are error-prone and subjected to both transition and transversion mutations. The transition mutations (purine to purine or pyramidine to pyramidine) are incorporated at a frequency of ~10^{-5} and transversion mutations (pyramidine to purine or *vice versa*) are incorporated at a frequency of 10^{-6}-10^{-7}. The transition mutations are generated at higher frequency than transversion mutations. However, the transition mutations are less likely to result in amino acid substitutions due to "wobble" mechanism, and are therefore, more likely to persist as "*silent substitutions*" as single nucleotide polymorphisms (SNPs) in viral populations. However, RdRps from SARS-CoVs are associated with a proofreading mechanism and therefore, the error rates are very minimum in these viruses [6].

1.4 Basic Steps Involved in the Lifecycle of ssRNA Viruses

As viruses are non-living entities, they enter into other compatible living cells and lead a parasitic life. They are ubiquitous in nature and successfully infect all forms of life, viz. bacteria, fungi, plant and animals. Even though they infect all forms of life, they are very specific to the species they infect and hence never infect other species under normal circumstances. To survive and complete their lifecycle, they have to find the right kind of host cells which possess specific receptor(s) for them to bind and enter the cells. Generally, the viral envelope glycoprotein usually binds to the host cell receptor(s) and enters the cell through a process called endocytosis. Once they enter the cell, the fusion of the virus membrane with the host endosomal membrane releases the viral genome into the cytoplasm of host cells.

After successful entry into the host cells, the viral genome acts like host mRNA and is translated to produce the viral proteins. The translated viral proteins are of two types, viz. the structural and non-structural proteins (NSPs). They involve both in the replication of the viral genome and subsequent assembly into new viral particles and their release. In CoVs, the NSPs are first produced as a polyprotein and then cleaved by the viral endoproteases embedded on the polyprotein itself into functional proteins. One of the NSPs, viz. the RdRp makes large number of copies of the viral RNAs for viral assembly with the structural proteins. Replication takes place in the cytoplasm at the surface of endosomes. As the first step, a double stranded RNA (dsRNA) intermediate is formed from the genomic single stranded RNA. Now the dsRNA is used for providing large number of (+) ssRNA genomes. In the next step, the nuclear capsid protein covers the genomic RNAs and gets enveloped in the cytoplasm. In the final step, the new viral particles exit the cells by the budding process at the plasma membrane. Interestingly, these RNA viruses except the retro viruses complete their entire lifecycle within the cytoplasm itself. The characteristic viral reproduction number, R_0, is the average number of people each person with a disease can infect and it varies from virus to virus.

1.5 Nidoviruses

Viruses are broadly classified into either RNA viruses or DNA viruses depending on whether their genome is made up of an RNA or a DNA, respectively. Most of the viruses in nature are RNA viruses and use RNA genomes. For different types of RNA viruses and their classifications refer to Palanivelu [7]. The coronaviruses (CoVs) belong to Nidoviruses. The Nidoviruses are enveloped viruses with a positive, single-stranded RNA genome. The CoVs which belong to the Nidoviruses are known to have the largest genomes for RNA viruses, with their genome sizes ranging from 26 to 32 kb in length. The CoVs belong to the order *Nidovirales*, family *Coronaviridae* and subfamily *Coronavirinae*. (The order *Nidovirales* derives its name from the presence of nested 3′ mRNAs as *Nido* is a Latin word for “nest”). The CoVs which belong to the family *Coronaviridae* are host specific and primarily cause enzootic infections in birds and mammals. However, in the last few decades, they have shown to be capable of infecting humans as well. The outbreak of SARS-CoV-1 in 2003, MERS-CoV in 2013 and more

recently the SARS-CoV-2 in 2020 has demonstrated their adaptability and lethality when they cross the species barrier and infect humans. Based on genetic and antigenic criteria, the CoVs have been classified into four groups: α-, β-, γ- and δ-CoVs. It is interesting to note that most of the CoVs are not dangerous to humans. For example, out of 7 family members, only 3 (SARS-CoV-1, MERS-CoV and SARS-CoV-2) caused SARS in epidemic and pandemic proportions and all the three belong to β genus. The other 4 human coronaviruses (HCoV-OC43 (β), HCoV-HKU1 (β), HCoV-229E (α), and HCoV-NL63 (α)) cause most of the colds during the year but they are not a serious threat to healthy individuals. In this communication, only the SARS and SARS-related CoVs and their replication machinery with special reference to SARS-CoV-2 are discussed.

1.6 Genome Organization of SARS-CoVs

1. The genome size of CoVs ranges from 26 to 32 kb. The genome contains a 5′ cap structure and a 3′ poly-(A) tail, allowing it directly to act as an mRNA for translation in host cells. They are enveloped viruses and the RNA genome is covered by a nuclear capsid protein with helical symmetry. Coronaviruses are roughly spherical with surface proteins, viz. the spike, envelope and membrane proteins projecting from the membrane.
2. The viral genome is a non-segmented, positive-sense, single-stranded RNA.
3. The 5′ end of the genome invariably contains an untranslated region (5'-UTR) which is known to form multiple stem-loop structures required for transcription and replication of the RNA. In addition to, in all Nidoviruses the genomic and sub-genomic RNAs share a common 5' leader sequence.
4. The 3' end of the genome also contains an untranslated region (3′-UTR) which is also known to form RNA secondary structures required for replication and synthesis of viral RNA.
5. The CoV genomes are organized as: 5′Cap—5'-UTR→ORFs 1a and 1b→S (Spike protein)→E (Envelope protein)→M (Membrane protein)→NCP (nuclear capsid protein) →3′-UTR-pol-(A) tail with accessory protein genes interspersed within the structural protein genes at the 3′ end of the genome (Fig. 2).
6. The main ORFs 1a and 1b encode the NSPs and occupy two-thirds, ~20 kb of the viral genome, as opposed to the structural and accessory proteins, which take up only ~10 kb of the genome.
7. The CoV genomes contain two overlapping reading frames ORF 1a and 1b, which direct the synthesis of two precursor polyproteins. (The expression of ORF1b occurs by a ribosomal frame-shift mechanism as the ORF1a/ORF1b overlapping nucleotide sequence is shown to be capable of inducing ribosomal frame-shift both *in vitro* and *in vivo*. In other words, the 1b is translated by a −1 frame-shift mechanism involving a pseudoknot structure). ORFs 1a and 1b encode 16 non-structural proteins and are highly conserved throughout CoVs.
8. The ORFs 1a and 1b NSPs are produced as a polyprotein and the polyprotein self-cleaves to form 16 NSPs (NSP1–NSP16) [8].
9. The sub-genomic RNA produce rest of the 4 essential structural proteins, viz. S, E, M, and NCP. Interspersed between these main structural gene reading frames, are the internal reading frames for the accessory proteins. The number of accessory proteins and their function is unique depending on the type of CoV (Fig. 2). The structural proteins like the NSPs also share common features in all CoVs [8].
10. The accessory proteins are almost exclusively nonessential for replication of the genome but some of them have been shown to be required for viral pathogenesis [9].

ORFs 1a and 1b genes encode 16 non-structural proteins (NSP1–11 and NSP12-16, respectively).

The structural genes encode the structural proteins, viz., S, E, M and NCP.

The accessory genes (shown in red) are unique to different CoVs in terms of number, organization, sequence, and function [10]. Some of them are embedded in the main structural protein genes as internal ORFs.

Typical UTRs for β-coronaviruses are 265 nt at the 5′ end and 229 nt at the 3′ end of the RNA. In SARS-CoV-2, the 5'- and 3'-UTRs are 265 and 358 nucleotides long, respectively [8].

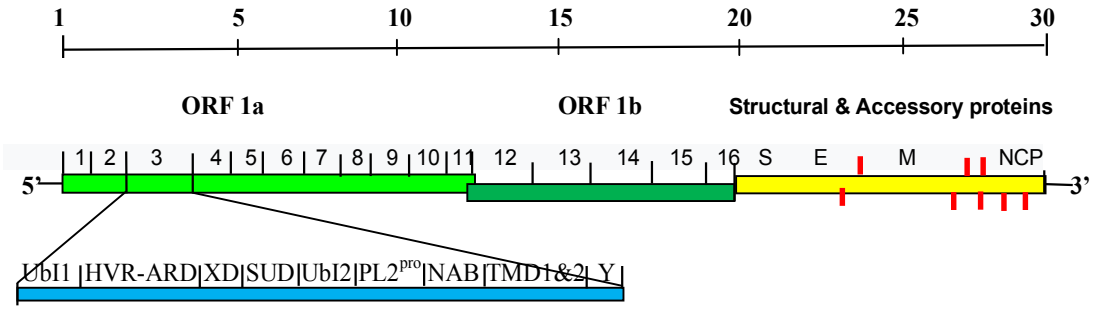

Fig. 2. A schematic diagram of the genome organization of COVs

1.7 Production of Polyprotein and Polyprotein Processing in CoVs

The ORFs 1a and 1b are translated as polyproteins and then processed by the viral proteases. This strategy is used by many viruses to ensure production of all the essential proteins for completing the lifecycle at one-go. These polyproteins thus, produced are subsequently cleaved into individual proteins by the viral proteases embedded on the polyprotein itself. For polyprotein processing, the CoVs harbour two or three protease genes (usually of different types) in its genome to cleave the polyprotein (CoVs do not depend on the host proteases for their polyprotein processing). These are the papain-like proteases (PLpro), encoded within *nsp*3, and a serine-type of protease, (Mpro) also known as the main protease encoded by *nsp*5. Most of the CoVs encode two PLpros within *nsp*3, except the β-CoVs like SARS-CoVs and MERS-CoV, where they produce only one PLpro. The PLpro cleaves NSP1/2, NSP2/3, and NSP3/4 boundaries, whereas the Mpro cleaves the rest. The replication–transcription is accomplished by the membrane-bound replicase–transcriptase complex (RTC). The RTC is composed of by a few proteins from the processed polyproteins. The NSPs and their proposed functions are discussed elsewhere.

1.8 Characteristics of the NSPs, Structural and Accessory Proteins

1.8.1 NSPs and their functions

ORF-1a: (NSPs1-11)

NSP1*: Suppresses antiviral host responses by blocking host-cell translation and promoting cellular mRNA degradation, which results in suppressing the host innate immune responses. This is achieved by the NSP1 by associating strongly with 40S ribosomal subunit and shutting-down the host mRNA translation both *in vitro* and *in vivo.* Furthermore, its binding to the ribosome also leads to endonucleolytic cleavage and subsequent degradation of host mRNAs. (However, interactions between the NSP1 and a conserved region in the 5'-UTR of viral mRNA prevent shutdown of viral protein expression through an unknown mechanism) [11].

NSP2: Unknown Function

NSP3: Multi Domain Region*

NSP4 and NSP6: are transmembrane proteins which form double-membrane vesicles (DMVs) from the virus-induced network of host endoplasmic reticulum membranes; also known as viral replication organelle. (The DMVs offer favourable microenvironment for viral RNA synthesis and also protect the double-stranded replicative forms from innate immune sensors that are activated by double-stranded RNAs). The newly synthesized RNAs, single-stranded forms, are exported to the cytosol for translation and viral genome assembly through the molecular pores of DMVs [12].

NSP5: Mpro (Main protease or CL protease)

NSP6: Type I IFN antagonist; suppresses the induction of both IFN as well as the IFN signaling pathways.

NSP7: Complex with NSP8 Primase for functional RTC

NSP8: Primase (NSP7 and NSP8 are known to make a hexadecameric complex that function as an effective primase to synthesize primers for the NSP12 RNA polymerase).

NSP9: ssRNA Binding Protein

NSP10: Stimulates NSP14 and NSP16 activities and thus, involves in cap and 2'-*O*-methylaton activities

NSP11: A short peptide at the end of 1a [8].

*The **multi-domain NSP3** is the largest protein encoded by the CoV genome, with an average molecular mass of ~200 kDa. At least, nine different domains are characterized in the NSP3 of all known CoVs. These regions are as follows: (Fig.1).
UBI1, Ubiquitin-like 1 domain; **HVR-ARD**, hyper variable Acid (Glu)-Rich Domain; **X** domain; **SUD**, SARS-CoV Unique domain; **UBI-2**, Ubiquitin-like 2 domain; **PL2pro**, Papain-like protease domain (releases NSP1, 2 and 3); **NAB**, Nucleic acid binding zinc-finger domain; **TMD**, Transmembrane domains1 & 2; **Y,** Y-Domain, highly conserved in all Nidoviruses [13].

ORF-1b: (NSPs 12-16)

NSP12: RdRp, activated by NSP8 and NSP7 complex of ORF1a
NSP13: Helicase (RNA unwinding enzyme with 5→3' polarity). NSP12 and NSP13 are expressed as part of two different cleavage products residing next to each other in ORF-1b in all Nidoviruses. The helicase is multifunctional and is implicated in transcription, replication, translation and virion biogenesis.
NSP14: ExoN, is a multifunctional protein. It possesses both N7-guanine methyltransferase (MTase) and exonuclease proofreading activities. The N7-guanine MTase adds 5'-Cap to the viral mRNAs and the 3'→5' exonuclease acts as the proof-reader for RdRp.
NSP15*: NendoU (a Uridine-specific endoribonuclease, chops-up leftover viral RNAs so that to hide it from the infected cell's antiviral defenses*).

*Type-I interferon I (IFN-1) induction and signaling represents one of the major innate antiviral defense pathways in humans, ultimately leading to the induction of several hundred antiviral interferon-stimulated genes (ISGs). SARS-CoV infections are first sensed by retinoic acid-inducible gene-I (RIG-I). The RIG-1 is nothing but a cytosolic pattern recognition receptor (PRR) which is responsible for the activation of the IFN-1 mediated host defense system [11,14].

(Interestingly, the ribonucleases, NSP14 and NSP15 are unique to the *Nidovirales* and are hence considered as genetic markers for these viruses).

NSP16: a 2'-*O*-ribose methyltransferase, essential for 5'-end cap modification to evade host immune system and also implicated in translation efficiency of viral mRNAs [8].

1.8.2 Structural proteins and their functions in CoVs

The viral envelope is made up of a lipid bilayer in which the surface proteins S, M and E are anchored in the ratio of E:S:M ~ 1:20:300.

The spike protein (~150 kDa, 1273 amino acids), is a heavily glycosylated protein (mainly with N-linked glycosylations). It assembles into a homotrimer and forms the distinctive spike-like structures on the surface of the virus (Analyzed in detail in my earlier communication [7].
The membrane protein (~25 kDa, 222 amino acids) is the most abundant structural protein found on the viral envelope and is also glycosylated. It is a smaller protein with three transmembrane domains and plays a central role in virus morphogenesis and assembly via its interactions with other viral proteins. It is also known to involve in the budding process of CoVs during viral exit from the host cells. It is interesting to note that the CoV-2 sequence shows remarkable similarity (98% identity) to the sequences from bat and pangolin isolates [15].
The envelope protein (~8 kDa, 75 amino acids) is the smallest one and also found in small quantities in the viral envelope. The E protein is highly conserved in CoVs. It is known to facilitate the viral assembly and release. It self-assembles and functions as a viroporins, forming pentameric protein-lipid pores on host membranes to allow ion transport. It also plays a role in the induction of apoptosis. Interestingly, sequence comparisons show that the CoV-2 envelope protein is almost identical to the sequences from bat and pangolin CoVs [15].
The nuclear-capsid protein (~46 kDa, 419 amino acids) forms the nuclear capsid. It oligomerizes to form a closed capsule-like structure, inside which the genome is securely packed. The nuclear capsid binds the viral genome as beads-on-a-string type conformation. It is composed of two separate domains, an N-terminal domain and a C-terminal domain which are separated by a linker region. The NCP is also heavily phosphorylated, and phosphorylation of NCP

has been suggested to trigger a structural change enhancing its affinity for viral versus nonviral RNAs. The usual N-terminal Met is removed and the new Ser at the N-terminal is acetylated. It plays a fundamental role during virion assembly through its interactions with the viral genome and the membrane protein M. It also plays an important role in enhancing the efficiency of subgenomic viral RNA transcription as well as viral replication. NCP is found to be the major antigen recognized by convalescent antisera, suggesting its potential as a diagnostic tool or for vaccine development [10].

1.8.3 Accessory proteins (APs): Their organization and functions in SARS-CoVs

The SARS-CoV-2 produces as many as 27 proteins/peptides out of which 16 of them are NSPs, 5 of them are structural proteins and 7 of them are accessory proteins (APs). The number and their arrangements of APs differ in different groups of CoVs like α, β, γ and δ. The APs of the three of the β-CoVs which caused the pandemics are only shown here (Fig. 3). Functions of the APs are outlined below [8].

The **ORF3a AP** makes a hole in the membrane of infected cells to make it easier for the new viruses to escape. It also triggers inflammation, one of the most dangerous symptoms of COVID-19.

The **ORF3b AP** is a novel short putative protein with 4 helices. The CoV-2 AP exhibits no homology to the existing SARS-CoVs or SARS-related-CoVs. However, the ORF3b still possessed Interferon (IFN) antagonist and Interferon Regulating transcription Factor-3 modulating activities, suggesting a potential link between SARS-related-CoVs from bats. [8].

The **ORF6 AP** blocks the signals send out to the immune system by the infected cells.

The **ORF7a AP** might be used to cut-down an infected cell's supply of tetherin (viral trapping mechanism of the host cells), allowing more of the viruses to escape. It also contributes to viral pathogenesis *in vivo* by inducing inflammatory responses (cytokine storm) in SARS and may also contribute to viral-induced caspase-3-dependent apoptosis.

The **ORF7b AP** is ~5 kDa integral membrane protein, localized to the Golgi compartment in SARS-CoV-infected cells.

The **ORF8 AP**'s function is not clear yet.

The **ORF9b AP** blocks IFNs, the body's key molecules which involve in first-line of defense against viruses.

The **ORF9c** overlaps the same stretch of ORF9b RNA, but with no known function [16].

1.8.4 Basic Structural Features of RdRps in CoVs

In general, the viral RdRps are structurally and functionally similar including that of the SARS-CoVs. For example, all of them exhibit a characteristic cupped right hand with three major domains, viz. Fingers, Palm and Thumb, which is similar to DNA polymerases and SSU RNA polymerases. In CoV-2, amino acids from 398 to 580 form the fingers domain, 581-812 form the palm domain including an interface of ~60 amino acids and 813-932 form the thumb domain. Amino acids 1-250 of the N-terminal region form the Nidovirus RdRp-Associated Nucleotidyltransferase (NiRAN) domain. The NSP7 and NSP8 are also known to interact with the NiRAN domain. Fig. 4 shows the arrangement of the various domains of the SARS-CoV-2 RdRp.

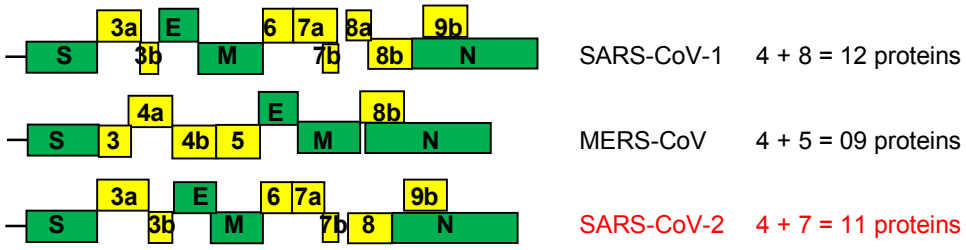

Fig. 3. Organization of APs in different SARS-CoVs

The four structural proteins are shown as green boxes and the accessory proteins as yellow boxes. Upper ones show the overlapping / internal ORFs

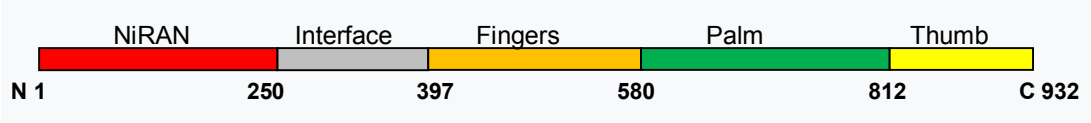

Fig. 4. A schematic diagram showing different domains of the SARS-CoV-2 RdRp

(The NiRAN domain is found in all COVs including the SARS-CoVs. The RdRp folds in such a way where the NiRAN domain encircles the active site). Adapted from Hillen et al [17]

1.8.5 RdRp form a Replicase–Transcriptase Complex (RTC)

For viruses, whose survival must rely on the efficient production of many progeny copies, stability, velocity and fidelity of replication are essential, and these features are constantly evolved and perfected during prolonged evolutionary times. The lifecycle of RNA viruses mainly depends on the successful transcription and replication to make all the essential structural and NSPs and viral genomes to infect more and more cells. Interestingly, in RNA viruses both replication and transcription are performed by the same enzyme, viz. the RdRp and both the processes are biochemically equivalent, often catalyzed by a multienzyme complex known as RTC. In CoVs, the replication complex is yet to be characterized in sufficient details both *in vitro* and *in vivo.* The multienzyme RTC is known to be composed of the RdRp (NSP12), the primase complex (NSP7 and NSP8) and other associated enzymes like helicase, proofreading exonuclease, methyltransferase, etc. The key component of the RTC complex is the RdRp, which catalyzes the synthesis of viral RNA and thus, plays a central role both in the replication and transcription cycle of all CoVs. Using biochemical assays and reverse genetics, Subissia et al [18] have studied the important conserved residues in NSP7, NSP8 and NSP12. They found that even though several NSP7 mutations affected virus replication to a limited extent, the two NSP8 mutations involving residues P^{183} and R^{190} were found to be essential for interactions with NSP12. Furthermore, a third mutation involving K^{58} was found to be critical for the interaction of the polymerase complex with RNA. All three mutations were found to lethal to the virus [18]. Besides, without a loss of processivity, the NSP7/NSP8/NSP12 complex can associate with NSP14, a bifunctional enzyme, possessing both the 3′-5′ exoribonuclease and RNA capping N7-guanine methyltransferase activities which are essential for replication fidelity and 5′-RNA capping [18-20].

1.8.6 RdRps and direct-acting antiviral drugs

RNA viruses, except the retro viruses, complete their entire life cycle within the cytoplasm itself and are able to exploit the host machineries for this purpose. At least 4 critical stages are found in their lifecycle. These 4 crucial stages in their lifecycle are the main targets to design direct-acting antiviral drugs (DAADs) to block each of these steps. At least 5 different types of DAADs are developed to block the viruses at each stage of their lifecycle [21], viz. i) Protease Inhibitors, ii) RdRp Inhibitors, iii) Interferon-Resisting Protein Inhibitors, iv) Glycosylation Inhibitors, v) Antioxidant Inhibitors. Only the inhibitors for the RdRp are elaborated further. As discussed elsewhere, the crucial step in the multiplication of these viruses is the transcription and replication processes which are essentially executed by the enzyme RdRp. A substantial obstacle for the discovery, optimization, and comprehensive evaluation of effective nucleotide based drugs for the SARS-CoV-2 is the lack of molecular details on the active sites, substrate recognition and catalytic mechanism during replication. Even though, no new inhibitors are designed for this CoV-2 RdRp, some of the broad-spectrum drugs designed for other RdRps are repurposed to block the RdRp of CoV-2 also. Some well-known examples are the broad-spectrum nucleoside analog inhibitors like Remdesivir (Ebola virus), Favipiravir (Influenza viruses), Ribavirin (Hepatitis C virus, HCV), Sofosbuvir (HCV), Galidesivir (HCV), Penciclovir (Herpesvirus), etc. These pro-drugs require efficient intracellular activation to their respective triphosphate forms to participate in the polymerization reactions. Therefore, it depends not only on their efficient conversion into their active forms in viral infected patients but also their availability in high concentrations to compete with the natural substrates. Secondly, the efficacy of these nucleoside analogues depends on their effective binding onto the active sites of CoV-2 RdRps for successful incorporation to block the RNA synthesis. Thirdly, unlike most of the RNA viruses, the CoVs also possess a proofreading activity.

Therefore, this communication, analyses the RTC of the SARS and SARS-related CoVs and their catalytic sites, metal binding sites and catalytic mechanism with special reference to CoV-2 RdRp.

2. MATERIALS AND METHODS

A large number of RdRp sequences of SARS and SARS-related CoVs are available in PUBMED and SWISS-PROT databases. For MSA analysis of various RdRps, the sequences were retrieved from SWISS-PROT and PUBMED sites and analyzed using Clustal Omega tool, an accurate, fast and widely accepted algorithm, available on their website. This communication presents the results obtained from the protein sequence analysis of these enzymes which are further supported by biochemical, SDM experimental data and X-ray crystallographic and cryo-EM studies of the SARS-CoV-2 RdRp and other related RdRps from positive and negative strand RNA viral pathogens available in the literature. The CoV-2 RdRp is used as the model enzyme for delineating the polymerization mechanism based on the above data.

3. RESULTS AND DISCUSSION

3.1 Analyses of Active Site Regions of RdRps of SARS-CoVs

Figure 5 shows the MSA analysis of the RdRps of SARS-related CoVs including SARS-CoV-2. The RNA polymerase sequences from different CoVs reveal conservation of large peptide regions among them. The MERS-CoV RdRp sequence varies in many places but completely conserved in the crucial catalytic regions. Besides, a three-amino acid deletion at the N-terminal region is seen in all except MERS-CoV RdRp (Fig. 5). The catalytic region and catalytic amino acid of the CoV-2 RdRps were arrived at from the MSA and cryo-EM data [17], and also from the data already available on other RdRps, DdRps and DNA polymerases [22-26]. The prokaryotic and eukaryotic multi-subunit (MSU) DdRp catalytic subunits invariably use the catalytic amino acid R followed by an invariant Y/FG (except human RNA polymerase where it uses a KG) whereas the single subunit (SSU) RNA polymerases and DNA polymerases mostly use the catalytic amino acid K followed by an invariant YG pair (Table 1). MSA analysis of the SARS and SARS-related CoVs uses an invariant K which is followed by the invariant YG pair where the invariant K could be possibly the catalytic amino acid as in DdDps (Table 1). The YG which is commonly found in all these nucleic acid polymerases is implicated in template binding [22-25]. Wang et al. [2] investigated the molecular basis of SARS-CoV-2 RNA replication by analyzing the cryo-EM structures of the stalled pre- and post-translocated polymerase complexes. They found in the pre-translocated complex, N496, K500, S501, N507, and K577 from the fingers domain and Y595 from the palm domain directly participate in stabilizing the phosphate backbone of the template. Interestingly, the same Y595 is in fact, part of the invariant YG pair (highlighted in yellow) found in all SARS-CoVs RdRp (Fig. 5) These findings were further corroborated by cryo-EM analysis of the RdRp of CoV-2 by Hillen et al [17], where they have also shown that the Y595 in the palm domain of the CoV-2 RdRp. However, in primases (NSP8) an invariant YA pair is identified as the probable template binding pair in the SARS-related CoVs and the K as the probable catalytic amino acid as in the RdRps (Fig. 6). The possible metal (Mg^{2+}) binding motifs in the RdRps are highlighted in light green. The Zn^{2+} binding motif (CxxCxxxxCxxxC-) highlighted in orange suggesting that like other RNA polymerases this enzyme also could be a Zn metalloenzyme.

One of the most conserved universal motifs of all the RdRps is the –GDD- and is shown to be essential for the catalytic function of RdRps [27,28]. This motif is also found (-G^{823}DD- in the thumb region) in CoV and CoV-related RdRps. In the secondary structure analysis of –GDD- motif, the G is found to involve in a tight turn and modifying this G→A by SDM experiments completely abolished the RdRp activity in Encephalomyocarditis virus [29]. The functional role of G in the –GDD- motif was further confirmed by similar SDM experiments in the replicase of a plus-strand RNA viruses like the Qβ bacteriophage and polio virus (PV). The substitution of the G of the –GDD- sequence by C. M, P or V residues inactivated the enzyme activity completely [30,31]. Additional conserved metal binding motif, viz. –ADD- is commonly found in many (+) strand viral pathogens. dengue viral RdRp crystals soaked in 0.2 M $MgCl_2$ clearly revealed the presence of one hydrated Mg^{2+} ion close to the expected catalytic position, hexa-coordinated by Asp-533 of -ADD^{533}- (motif A) via a water molecule and by Asp-664 of -GDD^{664}- (motif C). Interestingly, the –ADD- and –GDD- motifs are also completely observed in the RdRps of Yellow Fever, Japanese Encephalitis and West Nile viruses [32]. These two motifs (-

A/SDD- and –GDD-) are also found in SARS-CoV RdRps as well. The invariant –S^{759}DD- motif is located in the active site palm region of SARS-CoV-2 RdRp [17], which may be equivalent to the –ADD- motif of other viral RdRp as discussed above and participates in binding to the Mg^{2+} ion (Fig. 5).

Lehman et al. [33] have shown that the NiRAN domain which is unique to the CoVs is shown to exhibit nuceotidtyl-transferase activity [33]. Furthermore, they have also suggested K^{73} (marked in red) could be in the active site of the NiRAN domain as the K^{73}→A mutation dramatically reduced the viral titre. In addition to, K^{73} they also found R^{116}→A, K^{103}→A, D^{126}→A, D^{218}→A mutations (highlighted in green) reduced the viral titre markedly suggesting that they may also possibly in the active site region of the NiRAN domain. The D^{126} and D^{218} are found in catalytic regions by MSA analysis also. However, a template binding YA pair as in primases (NSP8) preceded by an invariant proton acceptor R and an NTP binding basic amino acid H at -5 are found in this domain and highlighted in yellow. Interestingly, an invariant YAN motif (highlighted in yellow) which is shown to be involved in NTP selection in CoV-2 RdRp by binding to the 2-OH of the NTP is also found. They [17,2] have also found this motif is completely conserved in all SARS-related CoV RdRps (except in MERS-CoV RdRp). In addition to, the Mg^{2+} binding –DxD- and -D^{218}XXD- motifs are also identified in this NIRAN domain by MSA analysis and are highlighted in green.

As mentioned in my earlier communication on spike proteins [7] the SARS-CoV-1 and civet-CoV exhibit very close sequence identity to each other in many places in RdRps and primases (highlighted in magenta), further confirming the observation that the civet-CoV could be a possible intermediate for SARS-CoV-1.

3.2 Analysis of the Primases of the RTC of SARS-CoVs

It has been shown that the RdRps of SARS-CoVs are a primer-dependent enzyme and it forms the holoenzyme by complexing with the primase subunit NSP8 and an associated protein NSP7 [8,17,19].

Figure 6 shows the MSA analysis of the primase enzyme in SARS and SARS-related CoVs. There is no universal –GDD- motif in primases as found in all (+) strand RdRps. However, other DxD type metal binding motifs which could bind the Mg^{2+} at the catalytic site is found and highlighted in green. Instead of a regular –YG- template binding pair, a YA pair is found (highlighted in yellow). Except in MERS-CoV RdRp, all the other SARS and SARS-related CoV primases show the –YA- pair with an S as –YAS- which is also found in SSU DdRps as -YGS-. [23]. In fact, the S in the –YGS- is shown to be essential for nucleotide discrimination in SSU RNA polymerases. In SSU RNA polymerases, the invariant S with the YG pair (e.g., T7 RNAP in Table 2) is shown to be important in nucleotide discrimination. It is interesting to note, that a single mutation modifying the S in –YGS- to A by a SDM experiment, the RNA polymerase behaved like a DNA polymerase, accepting dNTPs, suggesting that Ser–OH might involve in nucleotide discrimination by binding to the 2'-OH of the incoming NTPs [34,23]. The amino acid K in the invariant KL (marked in red) is predicted to play the catalytic role in the polymerization process as in RdRps. It is interesting to note that the invariant KL and YG pairs are common catalytic diads in eukaryotic DNA polymerases too (Table 1) [23].

The function of NSP8 as a primase in SARS-CoVs has been further proved unequivocally by Imbert et al [35]. They found that the primase synthesized *de novo* up to 6 nucleotides in a sequence-specific manner which could be subsequently used as a primer for RdRp for RNA synthesis [35,36]. It is interesting to note that Imbert et al. [35] have also shown that the NSP8 uniquely encode a Mn^{2+} dependent, rifampicin insensitive, second RdRp in SARS-CoVs but with relatively with low fidelity which closely agrees with the MSA findings (in fact, it lacks an NTP binding R at -5 from the catalytic pair KL, which is found in the high fidelity RdRps). They also suggested that the RdRps and primases in SARS-CoVs might have a possible common origin which is also supported by the MSA analysis (Table 1). The SARS-CoV primases revealed marked sequence preferences like cellular primases where RNA synthesis starts with a purine, viz. G or A. Furthermore, GTP analogues such as 3'-dGTP efficiently inhibited primase activity suggesting the primases as an attractive antiviral target for SARS-CoVs. The primase activity was most sensitive to replacement of either of the three residues, viz. K^{58}→A, K^{82}→A and S^{85}→A. Subissi et al. [37] have also shown by SDM experiments that K^{58} (marked in red) in the highly conserved block is crucial for the assembly of the primase with

RdRp as the conversion of K→A affects not only the assembly but also RNA synthesis [37]. It is known that NSP8 and NSP7 confer processivity to the RdRps and hence the mutation at K^{58} was found to be lethal for the SARS-CoV RdRps [37]. Subissi et al. [37] also found that the replacement of two more NSP8 residues (P^{183} and R^{190} at the C-terminal end, marked in red) were essential for interaction with RdRp (NSP12) and a third amino acid residue K^{58} was critical for the interaction of the polymerase complex and with RNA and therefore, all the three amino acid modifications were found to be lethal to the virus. Cryo-EM studies have shown that the CoV-2RdRp was bound to a heterodimer of NSP7 and NSP8 as well as to a second subunit of NSP8. The binding site for the NSP7-NSP8 heterodimer was found to be well conserved and overlapped with the conserved regions of polymerase functional domains (fingers and thumb) and all three subunits are essential for RNA synthesis [17].

CLUSTAL O (1.2.4) MSA of SARS-related CoVs' RdRps

```
MERS-CoV            -SKDSNFLNRVRGSIVNARIEPCSSGLSTDVVFRAFDICNYKAKVAGIGKYYKTNTCRFV   59
SARS-CoV-1          SADASTFLNRVCG-VSAARLTPCGTGTSTDVVYRAFDIY--NEKVAGFAKFLKTNCCRFQ   57
Civet-CoV           SADASTFFKRVCG-VSAARLTPCGTGTSTDVVYRAFDIY--NEKVAGFAKFLKTNCCRFQ   57
Pangolin-CoV        SADAQSFLNRVCG-VSAARLTPCGTGTSTDVVYRAFDIY--NDKVAGFAKFLKTNCCRFQ   57
SARS-CoV-2          SADAQSFLNRVCG-VSAARLTPCGTGTSTDVVYRAFDIY--NDKVAGFAKFLKTNCCRFQ   57
Wuhan_Std CoV-2     SADAQSFLNRVCG-VSAARLTPCGTGTSTDVVYRAFDIY--NDKVAGFAKFLKTNCCRFQ   57
Bat_RaTG13-CoV      SADAQSFLNRVCG-VSAARLTPCGTGTSTDVVYRAFDIY--NDKVAGFAKFLKTNCCRFQ   57
                     :. ..*::** * :  **: **.:* *****:*****    : ****:.*: *** ***

MERS-CoV            ELDDQGHHLDSYFVVKRHTMENYELEKHCYDLLRDCDAVAPHDFFIFDVDKVKTPHIVRQ   119
SARS-CoV-1          EKDEEGNLLDSYFVVKRHTMSNYQHEETIYNLVKDCPAVAVHDFFKFRVDGDMVPHISRQ   117  K73&103
Civet-CoV           EKDEEGNLLDSYFVVKRHTMSNYQHEETIYNLVKDCPAVAVHDFFKFRVDGDMVPHISRQ   117
Pangolin-CoV        EKDEDGNLIDSYFIVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQ   117
SARS-CoV-2          EKDEDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQ   117  R116
Wuhan_Std-CoV       EKDEDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQ   117
Bat_RaTG13-CoV      EKDEDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQ   117
                    * *::.: :****:*****:.**: *:  *:*::** *** **** * :* .*** **

MERS-CoV            RLTEYTMMDLVYALRHFDQN-SEVLKAILVKYGCCDVTYFENKLWFDFVENPSVIGVYHK   178
SARS-CoV-1          RLTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYAN   177
Civet-CoV           RLTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYAN   177
Pangolin-CoV        RLTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDEYFNKKDWYDFVENPDILRVYAN   177
SARS-CoV-2          RLTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYAN   177  D126
Wuhan_Std-CoV       RLTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYAN   177
Bat_RaTG13-CoV      RLTKYTMADLVYALRHFDEGNCDTLREILVTYNCCDDDYFNKKDWYDFVENPDILRVYAN   177
                    ***:*** ***********:. .:.*: ***.*.***  **::* *:******.:: ** :

MERS-CoV            LGERVRQAILNTVKFCDHMVKAGLVGVLTLDNQDLNGKWYDFGDFVITQPGSGVAIVDSY   238
SARS-CoV-1          LGERVRQSLLKTVQFCDAMRDAGIVGVLTLDNQDLNGNWYDFGDFVQVAPGCGVPIVDSY   237
Civet-CoV           LGERVRQSLLKTVQFCDAMRDAGIVGVLTLDNQDLNGNWYDFGDFVQVAPGCGVPIVDSY   237
Pangolin-CoV        LGERVRQALLKTVQFCDAMRDAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVVDSY   237
SARS-CoV-2          LGERVRQALLKTVQFCDAMRNAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVVDSY   237  D218
Wuhan_Std-CoV       LGERVRQALLKTVQFCDAMRNAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVVDSY   237
Bat_RaTG13-CoV      LGERVRQALLKTVQFCDAMRDAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPIVDSY   237
                    *******::*:**:*** * .**:*************:********:. . **.** :****

MERS-CoV            YSYLMPVLSMTDCLAAETHRDCDFNKPLIEWPLTEYDFTDYKVQLFEKYFKYWDQTYHAN   298
SARS-CoV-1          YSILMPILTLTRALAAESHMDADLAKPLIKWDLLKYDFTEERLCLFDRYFKYWDQTYHPN   297
Civet-CoV           YSILMPILTLTRALAAESHMDADLAKPLIKWDLLKYDFTEERLCLFDRYFKYWDQTYHPN   297
Pangolin-CoV        YSILMPILTLTRALTAESHVDTDLTKPYIKWDLLKYDFTEERLKLFDRYFKYWDQTYHPN   297
SARS-CoV-2          YSLLMPILTLTRALTAESHVDTDLTKPYIKWDLLKYDFTEERLKLFDRYFKYWDQTYHPN   297
Wuhan_Std-CoV       YSILMPILTLTRALTAESHVDTDLTKPYIKWDLLKYDFTEERLKLFDRYFKYWDQTYHPN   297
Bat_RaTG13-CoV      YSILMPILTLTRALTAESHVDTDLTKPYIKWDLLKYDFTEERLKLFDRYFKYWDQTYHPN   297
                    ** ***:*.:* .*:**:* * *: ** *:* * :****: :: **:*********** *

MERS-CoV            CVNCTDDRCVLHCANFNVLFAMTMPKTCFGPIVRKIFVDGVPFVVSCGYHYKELGLVMNM   358
SARS-CoV-1          CINCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHNQ   357
Civet-CoV           CINCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHNQ   357
Pangolin-CoV        CVNCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHNQ   357
SARS-CoV-2          CVNCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHNQ   357
Wuhan_Std-CoV       CVNCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHNQ   357
Bat_RaTG13-CoV      CVNCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHNQ   357
                    *:** ****:*********: .:* *.***:**************** ***::***:* *
```

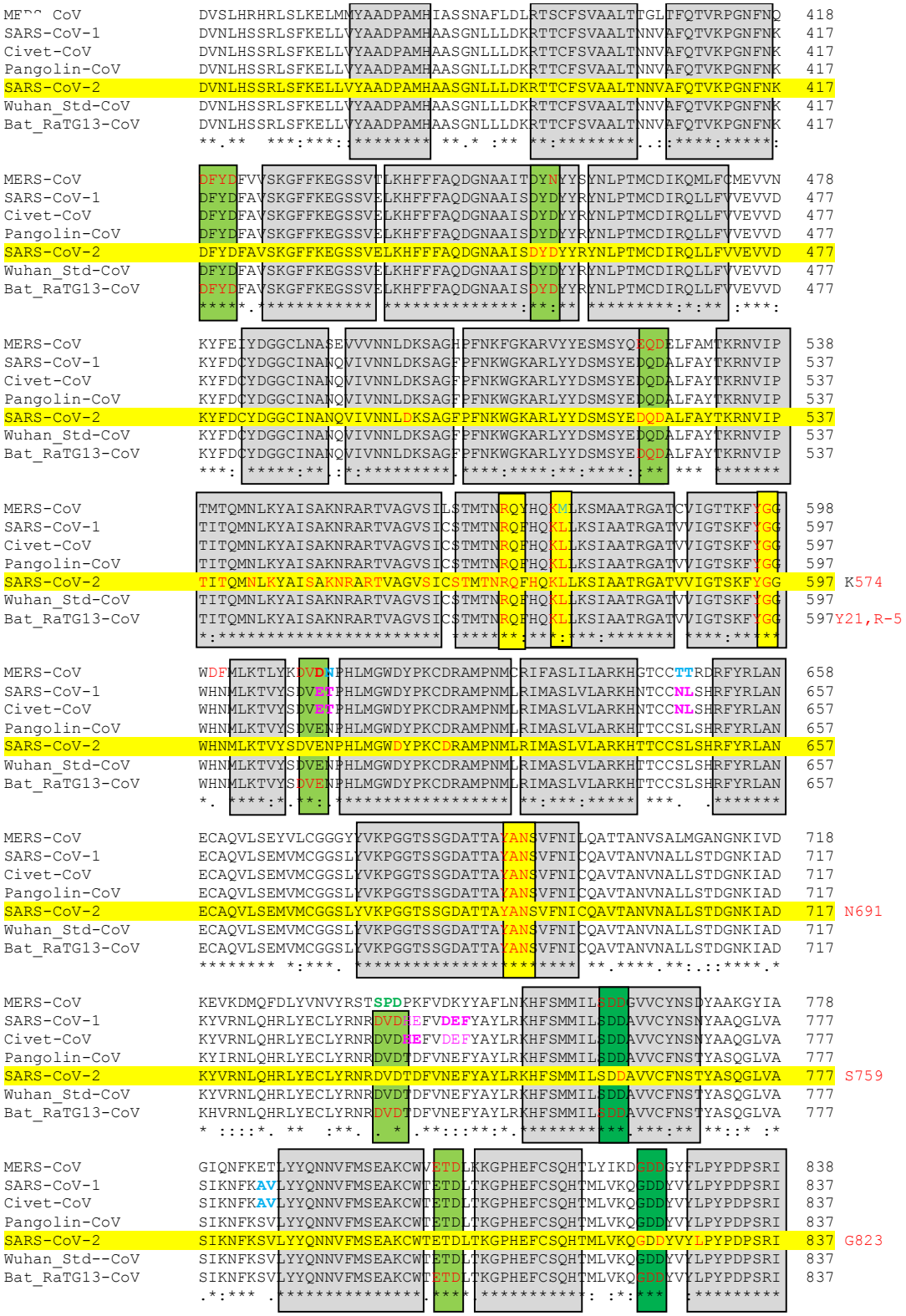
MERS-CoV DVSLHRHRLSLKELMNYAADPAMHIASSNAFLDLRTSCFSVAALTTGLTFQTVRPGNFNQ 418
SARS-CoV-1 DVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFNK 417
Civet-CoV DVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFNK 417
Pangolin-CoV DVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFNK 417
SARS-CoV-2 DVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFNK 417
Wuhan_Std-CoV DVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFNK 417
Bat_RaTG13-CoV DVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFNK 417
MERS-CoV DFYDFVVSKGFFKEGSSVTLKHFFFAQDGNAAITDYNYYSYNLPTMCDIKQMLFCMEVVN 478
SARS-CoV-1 DFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLLFVVEVVD 477
Civet-CoV DFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLLFVVEVVD 477
Pangolin-CoV DFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLLFVVEVVD 477
SARS-CoV-2 DFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLLFVVEVVD 477
Wuhan_Std-CoV DFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLLFVVEVVD 477
Bat_RaTG13-CoV DFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLLFVVEVVD 477
MERS-CoV KYFEIYDGGCLNASEVVVNNLDKSAGHPFNKFGKARVYYESMSYQEQDELFAMTKRNVIP 538
SARS-CoV-1 KYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALFAYTKRNVIP 537
Civet-CoV KYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALFAYTKRNVIP 537
Pangolin-CoV KYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALFAYTKRNVIP 537
SARS-CoV-2 KYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALFAYTKRNVIP 537
Wuhan_Std-CoV KYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALFAYTKRNVIP 537
Bat_RaTG13-CoV KYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALFAYTKRNVIP 537
MERS-CoV TMTQMNLKYAISAKNRARTVAGVSILSTMTNRQYHQKMLKSMAATRGATCVIGTTKFYGG 598
SARS-CoV-1 TITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYGG 597
Civet-CoV TITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYGG 597
Pangolin-CoV TITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYGG 597
SARS-CoV-2 TITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYGG 597 K574
Wuhan_Std-CoV TITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYGG 597
Bat_RaTG13-CoV TITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYGG 597Y21,R-5
MERS-CoV WDFMLKTLYKDVDNPHLMGWDYPKCDRAMPNMCRIFASLILARKHGTCCTTRDRFYRLAN 658
SARS-CoV-1 WHNMLKTVYSDVETPHLMGWDYPKCDRAMPNMLRIMASLVLARKHNTCCNLSHRFYRLAN 657
Civet-CoV WHNMLKTVYSDVETPHLMGWDYPKCDRAMPNMLRIMASLVLARKHNTCCNLSHRFYRLAN 657
Pangolin-CoV WHNMLKTVYSDVENPHLMGWDYPKCDRAMPNMLRIMASLVLARKHTTCCSLSHRFYRLAN 657
SARS-CoV-2 WHNMLKTVYSDVENPHLMGWDYPKCDRAMPNMLRIMASLVLARKHTTCCSLSHRFYRLAN 657
Wuhan_Std-CoV WHNMLKTVYSDVENPHLMGWDYPKCDRAMPNMLRIMASLVLARKHTTCCSLSHRFYRLAN 657
Bat_RaTG13-CoV WHNMLKTVYSDVENPHLMGWDYPKCDRAMPNMLRIMASLVLARKHTTCCSLSHRFYRLAN 657
MERS-CoV ECAQVLSEYVLCGGGYYVKPGGTSSGDATTAYANSVFNILQATTANVSALMGANGNKIVD 718
SARS-CoV-1 ECAQVLSEMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQAVTANVNALLSTDGNKIAD 717
Civet-CoV ECAQVLSEMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQAVTANVNALLSTDGNKIAD 717
Pangolin-CoV ECAQVLSEMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQAVTANVNALLSTDGNKIAD 717
SARS-CoV-2 ECAQVLSEMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQAVTANVNALLSTDGNKIAD 717 N691
Wuhan_Std-CoV ECAQVLSEMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQAVTANVNALLSTDGNKIAD 717
Bat_RaTG13-CoV ECAQVLSEMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQAVTANVNALLSTDGNKIAD 717
MERS-CoV KEVKDMQFDLYVNVYRSTSPDPKFVDKYYAFLNKHFSMMILSDDGVVCYNSDYAAKGYIA 778
SARS-CoV-1 KYVRNLQHRLYECLYRNRDVDHEFVDEFYAYLRKHFSMMILSDDAVVCYNSNYAAQGLVA 777
Civet-CoV KYVRNLQHRLYECLYRNRDVDHEFVDEFYAYLRKHFSMMILSDDAVVCYNSNYAAQGLVA 777
Pangolin-CoV KYIRNLQHRLYECLYRNRDVDTDFVNEFYAYLRKHFSMMILSDDAVVCFNSTYASQGLVA 777
SARS-CoV-2 KYVRNLQHRLYECLYRNRDVDTDFVNEFYAYLRKHFSMMILSDDAVVCFNSTYASQGLVA 777 S759
Wuhan_Std-CoV KYVRNLQHRLYECLYRNRDVDTDFVNEFYAYLRKHFSMMILSDDAVVCFNSTYASQGLVA 777
Bat_RaTG13-CoV KHVRNLQHRLYECLYRNRDVDTDFVNEFYAYLRKHFSMMILSDDAVVCFNSTYASQGLVA 777
MERS-CoV GIQNFKETLYYQNNVFMSEAKCWVETDLKKGPHEFCSQHTLYIKDGDDGYFLPYPDPSRI 838
SARS-CoV-1 SIKNFKAVLYYQNNVFMSEAKCWTETDLTKGPHEFCSQHTMLVKQGDDYVYLPYPDPSRI 837
Civet-CoV SIKNFKAVLYYQNNVFMSEAKCWTETDLTKGPHEFCSQHTMLVKQGDDYVYLPYPDPSRI 837
Pangolin-CoV SIKNFKSVLYYQNNVFMSEAKCWTETDLTKGPHEFCSQHTMLVKQGDDYVYLPYPDPSRI 837
SARS-CoV-2 SIKNFKSVLYYQNNVFMSEAKCWTETDLTKGPHEFCSQHTMLVKQGDDYVYLPYPDPSRI 837 G823
Wuhan_Std--CoV SIKNFKSVLYYQNNVFMSEAKCWTETDLTKGPHEFCSQHTMLVKQGDDYVYLPYPDPSRI 837
Bat_RaTG13-CoV SIKNFKSVLYYQNNVFMSEAKCWTETDLTKGPHEFCSQHTMLVKQGDDYVYLPYPDPSRI 837

Fig. 5. MSA analysis of the RdRps of SARS and SARS-related CoVs

CLUSTAL O (1.2.4) MSA of primases (NSP8)

Fig. 6. MSA analysis of primases (NSP8) in SARS and SARS-related CoVs

Interestingly, in primases also the SARS-CoV-1 and palm civet-CoV-1 sequences are very close to each other (marked in magenta, Fig. 6) as found in the spike proteins. [7].

Fig. 7 shows the MSA analysis of the NSP7 which is found to be associated with the SARS-CoV RdRp-primase complex. Only the N-terminal region is highly conserved in all but the C-terminal region is also conserved in all except in MERS-CoV NSP7 subunit. The -LQ diad appears to be the protease cleavage point during the polyprotein processing and is completely conserved in all. As there are no template binding YG or YA pairs as seen in the RdRps and primases, this subunit may possibly play only supportive role in the replication process. Its purely supportive role has been further proved by

mutational analysis too. For example, Subissi et al. [37] have shown that several mutations in the NSP7 subunit did not affect the viral replication to any significant level.

3.3 Analyses of Active Site Amino Acids in Various Domains of RdRps of RNA Viruses

Unlike other viral RdRps, RdRps of CoVs are composed of four distinct domains, viz. NiRAN, Fingers Palm, and Thumb. The polymerase domains, viz. fingers, palm and thumb are made up of three sets of highly/completely conserved amino acid(s)/motif(s) which perform distinct but highly coordinated functions during polymerization process [22,23]. The viral RdRps from both (–) and (+) strand viral pathogens are extensively studied by various investigators as they are the primary drug target to control the viral multiplication [37].

Cryo-EM analysis of the CoV-2 RdRp with 932 amino acids, has shown that the amino acids from 1-397 form the NiRAN domain including an interface and the rest of the protein from 398-932 form the RdRp. The NiRAN domain from the N-terminal is followed by an interface which connects the RdRp [17]. In RdRp, approximately amino acids from 398-580 form the fingers, 581-812 form the palm including an interface of ~60 amino acids and 813-932 form the thumb domains [17] (Fig. 8). All these domains are interconnected for effective coordination, e.g., the finger domain loops interconnect the fingers with the thumb domain and the NTP and template entry channels meet at the palm domain. Possible catalytic amino acid motifs and catalytic amino acids of various domains of SARS-CoV-2 RdRp are presented in Fig. 8.

3.3.1 Analysis of the NiRAN domain

NiRAN, Nidovirus RdRp-Associated Nucleotidyltransferase domain is a unique domain found only in SARS and SARS-related CoVs RdRps and with no apparent viral or cellular homologues making it a second genetic marker for the order Nidovirales. It exhibits a dual specificity for UTP and GTP. Furthermore, it is shown to be an essential domain of all RdRps of Nidoviruses. An invariant K, i.e., the K^{73} in all SARS-CoV RdRps and K^{75} in MERS-CoV RdRp, was found in the conserved motif of this domain (marked in red) and its importance discussed elsewhere.

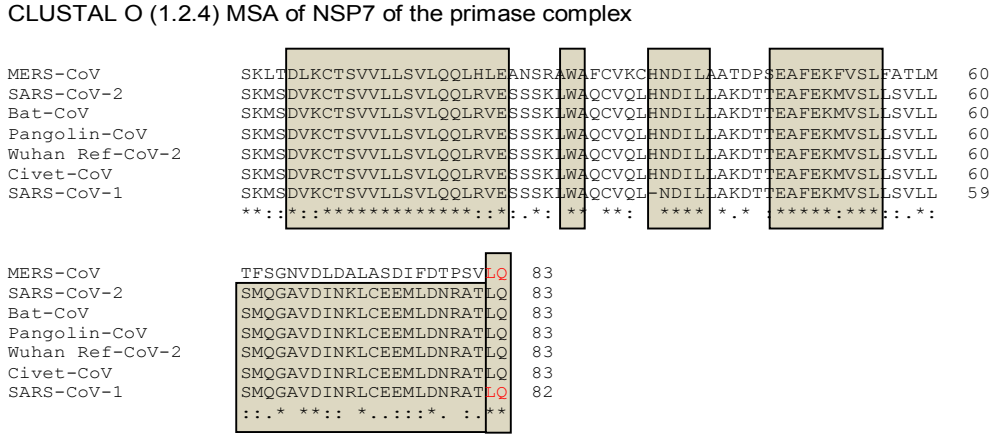

Fig. 7. MSA analysis of NSP7 in RdRp complex of SARS-related CoVs

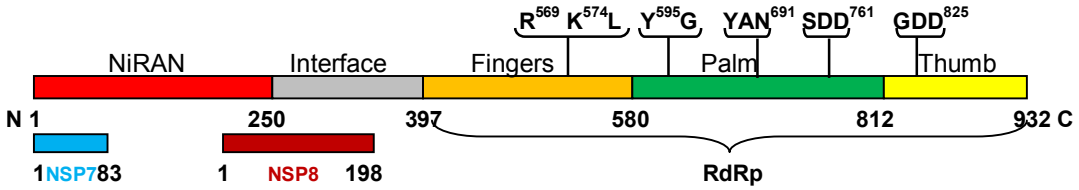

Fig. 8. Organization of different domains and their active site regions of the RdRp of SARS-CoV-2 and its interaction with the primer complex NSP7 and NSP8. Adapted from [17]

Replacement of the K^{73}→A by SDM experiments has shown that the viral RdRp (equine arteritis) lost its replication activity in cell culture studies. In addition to its role in replication, other important functions are also implicated for this domain, such as ATP-dependent RNA ligase, mRNA capping (guanylyltransferase), UTP/GTP binding and nucleotidylation, and protein-primed RNA synthesis, etc. [33]. Interestingly, the N-terminal residues of the polio viral RdRp also exhibited strong interaction with the C-terminal thumb region. The N-terminal residues are found to be essential for polymerase activity; deletion of Trp5 or of residues 1–6 completely inactivates the polio viral RdRp activity. In CoV-2 this region is shown to interact with the base of the polymerase palm and fingers domains [27].

3.3.2 Fingers domain

Amino acids from 398 to 580 form the fingers domain. It is a nucleotide interacting domain which primarily involves in NTP binding and selection in nucleic acid polymerases (Fig. 8). NTP and dNTP discrimination in nucleic acid polymerases is usually achieved by base pairing, base stacking, hydrogen bonding and by specific interaction(s) of the completely conserved basic amino acid (R/K/H) very close to the catalytic amino acid. Thus, NTP binding and selection is usually achieved by a basic amino acid which is placed very close to the catalytic amino acid in the downstream region. The invariant R/K/H is usually placed at -4 to -6 from the catalytic amino acid K/R and is also shown to be crucial for DNA and RNA polymerases (Table 1) [22,23]. Furthermore, this domain also harbours the catalytic amino acid, an invariant K/R in the catalytic region. Moreover, this domain is made up of many highly conserved, positively-charged basic amino acids and hydroxyl amino acids such as S/T in the highly conserved block preceding the NTP binding and catalytic amino acids (Fig. 5). Interestingly, this region is particularly devoid of any acidic amino acids like D or E. The Incoming NTPs would enter the active site through a tunnel at the back of this domain.

3.3.3 Palm domain

Amino acids from 581 to 812 including an interface region of ~60 amino acids form the palm domain (Fig. 8). This domain primarily contains the template binding YG pair and a nucleotide discrimination motif (highlighted in yellow). The MSA analysis shows that amino acids K and Y from K^{574}L to Y^{595}G pairs are likely to play a role in catalysis as reported for DdDps. In support of this MSA finding, the cryo-EM analysis also found that the proposed Y^{595} in the palm domain of CoV-2 RdRp. The cryo-EM study has also found the amino acids G^{590} and S^{592} in the conserved block preceding the YG pair [17]. These findings are further supported by Wang et al [2] who investigated the molecular basis of SARS-CoV-2 RNA replication by analyzing the cryo-EM structures of the stalled pre- and post-translocated polymerase complexes. They found in the pre-translocated complex N496, K500, S501, N507, and K577 from the fingers domain and Y^{595} from the palm domain directly participating in stabilizing the phosphate backbone of the template. An invariant template binding YG pair is also shown to be in the palm domain of DNA and RNA polymerases [22,23]. It is interesting to note that in all SARS and SARS-related CoVs the completely conserved template binding YG pair (highlighted in yellow) is found in a completely conserved block (Fig. 5). Thus, the invariant YG pair appears to be specific not only for the nucleic acid polymerases using DNA as the template (including the viral type SSU and prokaryotic and eukaryotic MSU DdRps (Table 1) but also RNA as template (RdRps).

Not only the Y but also the G in the YG pair was found to be essential for polymerase function. A point mutation G^{294}→A in the palm domain of the RdRp of the encephalomyocarditis virus (corresponds to the amino acid is G^{767} in the YG pair in *E. coli* DNA polymerase I) completely lost the activity. Another point mutation at Y^{293}→T of the YG pair in the same viral RdRp (corresponds to Y^{766} of the YG pair in *E. coli* DNA polymerase I which involves in the dNTP binding at 3-end of primer) showed only 6% activity [29].

In addition to, Hillen et al. [17] [2020] have found Y^{689} and N^{691} in the highly conserved –YAN- motif (highlighted in yellow) is also in the active site region of the palm domain. They found D^{760}, D^{761} of the –SDD- motif and N^{691} of the –YAN- motif in the active site. D^{760}→A mutation was incapable of RNA synthesis even in the presence of NSP7 and NSP8. Therefore, it was proposed that both Y^{689} and N^{691} (of the YAN motif) found at the active site, to recognize the 2'-OH group of the incoming NTP and thus, involve in the nucleotide

Table 1. Amino acids around the catalytic amino acid K/R and the YG/FG pair in DNA/RNA polymerases and SARS-CoV RdRps

Polymerase Type	Catalytic Region
SSU RNA and DNA Polymerases	
Viral T7 SSU RNA pol	$_{-620}$WLAYGVTR^{-5}SVTKR^{1}SVMTLAY^{8}GS-
Viral SP6 SSU RNA Pol	$_{-612}$WDSIGITR^{-5}SLTKK^{1}PVMTLPY^{8}GS-
Mitochondrial SSU RNA pol (Yeast)	-TR^{-5}KVVKQ^{1}TVMTNVY^{8}GV--
Chloroplast SSU pol (ARATH)	-DR^{-5}KLVKQ^{1}TVMTSVY^{8}GV-
E. coli DNA pol I (SSU)	-QR^{-5}RSAKA^{1}INFGLIY^{8}GM-
Initiation Subunits of Prokaryotic and Eukaryotic MSU RNAPs	
E. coli MSU RNAP β subunit	$_{-539}$TR^{-8}ERAGFEVRD^{1}VHPTHY^{7}G^{8}RV$_{558-}$
S. cerevisiae MSU RNAP II Rpb2 subunit	$_{-851}$FR^{-5}SLFFRS^{1}YMDQEKKY^{9}GMSI$_{870-}$
Human MSU RNAP II Rpb2 subunit	$_{-806}$FR^{-5}SVFYRS^{1}YKEQESKK^{9}GFDQ$_{825-}$
Elongation Subunits of Prokaryotic and Eukaryotic MSU RNAPs	
E. coli MSU RNAP β' subunit	$_{-833}$NSV^{-6}DAVKVRS^{1}VVSC^{5}DTDFGVC^{12}AHC^{15}Y^{16}G^{17}RDL$_{861-}$
S. cerevisiae MSU RNAP II Rpb1 subunit	$_{-55}$DPR^{-6}LGSIDRN^{1}LKC^{4}QTC^{7}QEGMNEC^{14}PGHF^{18}G^{19}HI$_{84-}$
Human MSU RNAP II Rpb1 subunit	$_{-59}$DPR^{-6}QGVIERT^{1}GRC^{4}QTC^{7}AGNMTEC^{14}PGHF^{18}G^{19}HI$_{88-}$
SARS-Related CoVs*	
RdRPs (NSP12)	
SARS-CoV-1	-STMTNR^{-5}QFHQKL^{1}LKSIAATRGATVVIGTSKFY^{21}GG$_{597-}$
SARS-CoV-2	-STMTNR^{-5}QFHQKL^{1}LKSIAATRGATVVIGTSKFY^{21}GG$_{597-}$
MERS-CoV	-STMTNR^{-5}QYHQKM^{1}LKSMAATRGATCVIGTTKFY^{21}GG$_{598-}$
NiRAN Nucleotidyl-transferase	$_{-112}$PH^{-5}ISR^{-2}QRL^{1}TKYTMADLVY^{11}AL$_{131-}$
Primases (NSP8)*	
SARS-CoV-1	$_{-120}$IPLTTAAKL^{1}MVVVPDYGTYKNTCDGNTFTY^{22}ASALWE-
SARS-CoV-2	$_{-120}$IPLTTAAKL^{1}MVVIPDYNTYKNTCDGTTFTY^{22}ASALWE-
MERS-CoV	$_{-121}$IPLCASNKL^{1}RVVIPDFTVWNQVVTYPSLNY^{22}AGALWD-
Active Site Regions in Prokaryotic and Eukaryotic DNA Polymerases#	
T4 DNA polymerase	$_{-546}$ATLANTN^{-5}QLNRKI^{1}LINSLY^{7}GALGNIH-
P. furiosus DNA polymerase	$_{-477}$KILLDYR^{-5}QKAIKL^{1}LANSFY^{7}GYYGYAK-
Yeast alpha DNA polymerase	$_{-933}$RVQCDIR^{-5}QQALKL^{1}TANSMY^{7}GCLGYVN-
Human alpha DNA polymerase	$_{-939}$ILQYDIR^{-5}QKALKL^{1}TANSMY^{7}GCLGFSY-
Human Gamma DNA polymerase	$_{-917}$TTVGISR^{-4}EHAKI^{1}FNYGRIY^{8}GAGQPFAER-
Human Delta DNA polymerase (Catalytic subunit)	$_{-683}$RQVLDGR^{-5}QLALKV^{1}SANSVY^{7}GFTGAQV-

** Bioinformatics analysis and predictions of functionality to the completely conserved motifs is important for further validation by experiments. #Adapted from Palanivelu [22]*

discrimination by selecting NTPs and not dNTPs during RNA synthesis [17,2]. Therefore, the invariant YG pair and YAN motif found not only in CoV-2 RdRp but also in all SARS and SARS-related CoV RdRps are shown in nucleotide binding and selection in the proposed mechanism. Thus, the palm domain plays a role in template binding and could bind the negatively charged triphosphate backbone of the incoming NTP at position +1. Its triphosphate group is further stabilized by the catalytic Mg^{2+} that is coordinated by the palm domain.

3.3.4 Thumb domain

Amino acids from 813 to 932 in the C-terminal region form the thumb domain (Fig. 8). This domain is mostly formed by amino acids, viz. R836, K849, M855, E857, R858, S861, I864, D865, Y915. Interestingly, the stretch of amino acids from 840 to 872 forms the highly conserved region with–YLQYI^{888}- diad repeat and highly conserved -HMLDMYSVML- and -YWE-motifs in all SARS and SARS-related CoV RdRps (a minor exception was observed in MERS-CoV RdRp) (Fig. 8).

The thumb domain primarily involves in correctly positioning the NTPs for the electrophilic–nucleophilic attack during polymerization reactions using their invariant Ds by forming Mg^{2+}-NTP complex, e.g., two highly conserved Ds coordinate Mg^{2+} ions in the catalytic site in all DNA and RNA polymerases including the SARS-CoV RdRps (Fig. 5). The thumb domain of various nucleic acid polymerases has been extensively analyzed by X-ray crystallographic and SDM experiments. Biochemical, crystallographic and SDM analyses of many RdRps of (+) strand and (–) strand viruses have shown that the palm domain is mainly consisted of the crucial catalytic metal binding amino acids which usually bind a Mg^{2}+ ion. These studies have unequivocally proved the presence of an invariant –GDD- motif which is known as the universal motif in all (+) strand viral RdRps [28]. This motif is also found in all SARS and SARS-related CoV RdRps as well (Fig. 5).

Sankar and Porter [29] analyzed the universal metal binding –GDD- motif in encephalomyocarditis (EMC) virus by SDM experiments. (a (+) strand, non-enveloped RNA virus, belonging to Picornaviridae, is the causative agent of not only myocarditis and encephalitis, but also neurological diseases, reproductive disorders and diabetes in many mammalian species) They found that all three amino acids in the –GDD- motif are essential for the RdRp activity. For example, replacements of G^{332}→A; D^{333}→E D^{334}→E yielded 0%, 3% and 0% activities of the enzyme, respectively.

X-ray crystallographic analysis of the RdRp of dengue virus (a (+) strand RNA viral human pathogen) clearly revealed the presence of one hydrated Mg^{2+} ion close to the expected catalytic position, hexa-coordinated by Asp-533 (-ADD^{533}-) via a water molecule and by Asp-664 (-GDD^{664}-) [32]. Interestingly, they also found that the -ADD- and –GDD- motifs are completely conserved in other (+) strand RNA viral pathogens' RdRps like Yellow fever virus, (YFV), Japanese encephalitis virus (JEV), and West Nile virus (WNV).

The functional role of –GDD- in catalysis is further corroborated by SDM analysis of the replicase of a (+) strand RNA viruses, the Qβ bacteriophage and PV. The substitution of the G of the –GDD- sequence by A, S, P, M or V residues in Qβ (C, P, M and V in PV) completely abolished the replicase activity [30,31], suggesting the importance of this universal motif in all RdRps including bacteriophages. Interestingly, this metal binding motif is also conserved in most viral RdRps, e.g., 317 to 319 (-GDD-) in HCV 3Dpol, ns5b and 327 to 329 (-GDD-) in PV RdRp.

The same motif is found in all SARS and SARs-related CoV RdRps also (highlighted in dark green) (Fig. 5). In addition to this universal motif, Hillen el al and Wang et al [17,2] have also found another conserved motif the -SDD^{761}- in the active site of SARS-CoV-2 RdRp. Interestingly, this metal binding motif is common in (-) strand viral pathogens (Table 2). The completely conserved –GDD- and -SDD- motifs in SARS and SARS–related RdRps are highlighted in dark green (Fig. 5).

3.4 Analyses of Metal Binding Sites in RdRps of + and – Strand RNA Viral Pathogens

In order to find out whether both the types of viral pathogens use same type of metal biding sites, the RdRps of both (+) and (–) RNA viruses were analyzed by MSA analysis (this communication), and a consensus is arrived at using SDM experiments and X-ray crystallographic data.

Almost all the nucleic acid polymerases invariably are known to use two Ds for the binding of the catalytic Mg^{2+} ion (Table 2).The (+) strand viruses, including the CoVs, invariably use the two Ds possibly from one from the–GDD- motif and the other from the –SDD- motif. As already discussed Sankar and Porter [29] have shown the point mutations generated by SDM experiments in the metal binding motif -G^{332}DD- drastically affected the polymerization activity of RdRp of EMC virus. Their results showed that all three amino acids are important in maintaining the structure at the metal (Mg^{2+}) binding region (Table 2).

Not only the (+) strand RNA viruses but also the - strand RNA viral pathogens such as influenza, measles, mumps and respiratory syncytial viruses and Ebola virus (EBOV) also use RdRp for genome replication and utilize Mg^{2+} as a co-factor for replication. However, MSA (this communication), SDM experiments and mini-genome assays [38] have shown that -GDNQ- is an invariant motif in all the – strand RNA viruses like Ebola, Influenza, Rabies Measles

and Mumps viral RdRps (Table 2). Schmidt and Hoenen [38] have found by SDM experiments that the –GD^{742}NQ- → -GAAA- mutation in Ebola viral RdRp abolished the replication as well as transcription activities. In addition to, they found by MSA analysis that the (–) strand RNA viral pathogens possessed the invariant –GDNQ- in their RdRps indicating that this motif is absolutely essential for genome replication, transcription, or both. Tchesnokov et al. [39] found that the D^{742}→A mutant in the –GD^{742}NQ- motif of Ebola viral RdRp lacked the ability to extend the primer in the presence of Mg^{2+} and therefore, it is clear that the Ebola viral RdRp mutant D^{742}→A abolishes the catalytic Mg^{2+} binding activity and hence the enzyme activity too. In addition to this motif, the (–) strand viral pathogens also possess a regular -G/SDD- motif(s) as found in (+) strand viral RdRps (Table 2). It is clear from Table 2 that –GDD- and –GDN- are signature metal binding sites in (-) strand viral RdRps.

However, the MSU DdRps of both prokaryotic (eubacterial and chloroplasts) and eukaryotic organisms use a different metal binding sites like –E/DxD-, -NxEDS- (for initiation reaction) and -NADFDGD- (for elongation reaction) (Table 2) [24-26,40]. By electron paramagnetic resonance spectroscopy, flow-dialysis and transcription analysis, the D^{537} and D^{812} in bacteriophage T7 RNA polymerase were found to be as metal ion-binding sites and are found to be essential for the catalytic mechanism [41]. The metal binding sites of both DdRps and RdRps are shown in Table 2.

3.5 Mechanism of Action of the RdRp of SARS-CoV-2

It is interesting to note that all the observed structural features of DNA and RNA polymerases suggest not only a strong conservation of structure-function relationships among them, but also a very similar active site amino acids and catalytic mechanism [22-26]. A minimal number of steps involved in the catalytic cycle of RdRps are: NTP selection, Watson-Crick base pairing with the complementary nucleotide to the template, catalysis, pyrophosphate release and translocation. Catalysis is initiated by nucleophilic attack on the α-phosphate of the base-paired NTP on the primer 3′-OH leading to formation of the phosphodiester bond and subsequent release of the pyrophosphate. Based on all the data available on SARS-CoV-2, Fig. 9 describes a plausible mechanism of action for the SARS-CoV-2 RdRp.

Step 1. Enzyme with the NTP at the Entry Site: Template binding by Y^{595}-G pair, NTP binding and selection by the invariant amino acid(s) at the catalytic site. (The invariant R^{569} is shown to play the important role in NTP binding and selection as reported for other nucleic acid polymerases. The guanido group of R^{569} could form the necessary hydrogen bonding with the triphosphate group of the substrate in the ternary complex). Watson-Crick base pairing is performed of the incoming NTP with the template RNA. The catalytic site amino acid K^{574} is positioned for proton abstraction. The Mg^{2+} which is coordinated by the catalytic D^{760} and D^{824}, which shields the β- and γ- phosphates and exposes the α-phosphate of NTPs for the nucleophilic attack at the polymerizing site. The second set of amino acids in the active site, viz. Y^{689} and N^{691} in the completely conserved –YAN- motif in all SARS and SARS-related CoV RdRps further reinforce the nucleotide discrimination by stacking interactions and recognizing the 2'-OH in the ribose moiety the NTPs. The Y is suggested to involve in base stacking as well as hydrogen bonding 2'-OH of ribose moiety and N is shown to recognize the 2'-OH of the of ribose moiety by hydrogen bonding. Thus, recognition of the 2'-OH of ribose moiety by Y and N becomes the major fidelity checkpoint here. YAN/Q type of motif is also found in the RdRps of Dengue, Zika, and Yellow Fever viruses too (data no shown).

Step 2. Proton Abstraction and Nucleophilic Attack of the α-phosphate of NTP: Electronic transition at the active site for proton abstraction by the active site K and for the initiation of electrophilic and nucleophilic attack.

Step 3. Phosphodiester Bond Formation: Proton abstraction by the active site amino acid K and simultaneous formation of phosphodiester bond. K/H is also proposed in the catalytic site of RdRp of PV to protonate the PPi leaving group during catalysis [36].

Step 4. Inorganic Pyrophosphate Formation and Translocation of the Enzyme to Next Nucleotide: Proton transfer from the active site amino acid K and formation of inorganic pyrophosphate; restoration of the active site and translocation of the enzyme to next nucleotide (Fig. 9).

Table 2. Metal binding sites in DNA- and RNA-dependent RNA polymerases

Subunit (Organism)	Metal biding site(s)	Method and Reference
Eubacteria (MSU RNA pol)		
β- subunit (*E. coli*)	-^{671}LEHDDA/ -^{809}GYNFEDS*-(Mg^{2+})	By MSA [24]
β'- subunit, Elongation (*E. coli*)	-^{458}YNADFDGDQM - (Mg^{2+}) & -^{883}RS^{1}VVSC^{5}DTDFGVC^{12}AHC^{15}Y^{16}GR^{901}-(Zn^{2+})*	By X-ray crystallography [40]
Eukaryote (MSU RNA pol)		
Rpb2 subunit (*S. cerevisiae*)	-^{893}LDDDG^{897}- -^{832}GYNQED*S^{838}-(Mg^{2+})	By MSA [25]
Rpb1 subunit, Elongation (*S. cerevisiae*)	–^{478}YNADFDGDEM^{487}-(Mg^{2+}) & -^{56}PR^{-6}LGSIDRN^{1}LKC^{4}QTC^{7}QEGMNEC^{14}PGHF^{18}GH^{83}-(Zn^{2+})	By MSA [25]
Chloroplast (MSU RNA pol)		
β-subunit (*Zea mays*)	-IE^{543}HNDA- &-GYN^{674}FEDA-(Mg^{2+})	By MSA [26]
β'-subunit (*Zea mays*)	-FN^{487}ADFDGDQM - (Mg^{2+})	By MSA [26]
β''-subunit (*Zea* mays)	-R^{291}S^{1}PFTC^{5}RSTSWIC^{12}QLC^{15}Y^{16}G-(Zn^{2+})	By MSA [26]
Mitochondria (SSU RNA pol)		
Arabidopsis thaliana	-QD^{677}GSC- & -VH^{908}DSF-	By MSA [23]
Nicotiana sylvestris	-QD^{703}GSC- & -VH^{934}DSY-	By MSA [23]
S. *cerevisiae*	-QD^{945}GTC- & -VH^{1188}DSY-	By MSA [23]
Candida albicans	-QD^{881}GTC- & -VH^{1121}DSF-	By MSA [23]
Neurospora crassa	-QD^{901}GTC- & -VH^{1180}DSF-	By MSA [23]
Chloroplast (SSU RNA pol)		
Arabidopsis thaliana	-QD^{694}GSC- & -VH^{925}DSY-	
Brsassica napus	-QD^{680}GSC- & -VH^{912}DSY-	
Bacteriophages (SSU RNA pol)		
T7 phage	-FD^{537}GSC- & -HD^{812}SFG-	By EPR and FD [41]
Lambda Phage	-FD^{537}GSC- & -HD^{812}SFG-	[23]
RdRps (EC 2.7.7.48) of (+) Strand Viruses		
CoV-2	-LSDD^{761}A- & -QGDD^{825}Y-	By X-ray crystallography] [19]
CoV-1	-LSD^{760}DA-	By SDM [35]
Polio	-YGDD^{329}-	[19]
Hepatitis C	NGDD^{319}-	[19]
Dengue	-YADD^{533}TA- & -SGDD^{664}CV-	By X-ray crystallography 32]
Zika	-YADD^{535}TA- & -SGDD^{666}CV-	By seq..similarity
Yellow Fever	-YADD^{284}TA- & -SGDD^{417}CV-	This communication
Rubella	-QGDD^{667}M- & -LGDD^{26}A-	This communication
Encephalomyocarditis	-YG^{332}DDL-	By SDM [29]
RdRps (EC 2.7.7.48) of (-) Strand Viruses		
Ebola	-MG^{741}DNQC- & -HTSDD^{617}FG-	By SDM [38]
Ebola	-MGD^{742}NQC- & ,,	By SDM [39]
Influenza A	-TGD^{305}NTK- & -QSSDD^{446}FA-	By seq. similarity
Influenza B	-TGD^{305}NTK- & -QSSDD^{445}FA-	This communication
Influenza C	-TGD^{307}NSK- & -QSSDD^{451}FV-	This communication
Rabies	-QGD^{729}NQV- & -LESDD^{1589}NI-	This communication
Nipah	-QGD^{832}NES- & -IDSDD^{1525}NI-	This communication
Measles	-QGD^{773}NQT- & -AYGDD^{1256}DS- & -LIGDD^{1470}D- & -ESDED^{1598}V-	This communication
Mumps	-QGD^{779}NQA- & -SSGDD^{1195}KF-	This communication

**Possible metal binding sites arrived at by MSA and SDM*; pol., polymerase; seq., sequence; EPRS, Electron Paramagnetic Resonance Spectroscopy; FD, Flow Dialysis.*

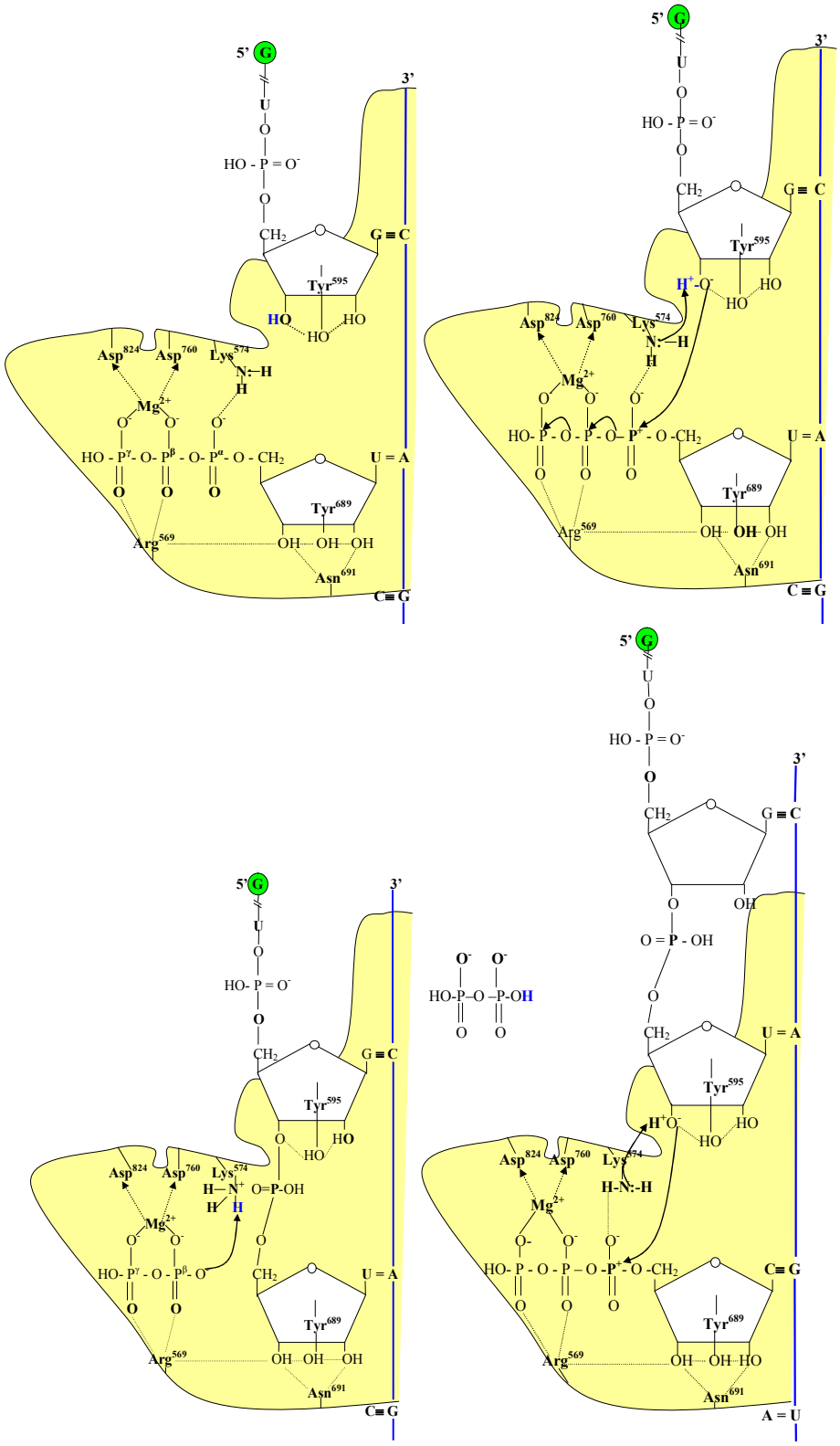

Fig. 9. Steps (1-4) involved in the proposed catalytic mechanism for CoV-2 RdRp

4. CONCLUSIONS

The MSA analysis of SARS and SARS-related CoV RdRps exhibited very high similarity among them. Though the MERS-CoV RdRp was deviating in many conserved motif regions, yet the active site regions and the catalytic amino acids are completely conserved. The RdRps of palm civet and SARS-CoV-1 exhibited very high similarity among themselves suggesting the possibility of palm civet as the possible intermediate. The universal metal binding motif in RdRps, viz. –GDD- and an additional metal binding motif –SDD- were found in all these SARS-CoV RNA polymerases. Active site regions and metal binding sites suggest that the RdRps should have evolved from DdDps (or *vice-versa*) rather than DdRps. The MSA analysis along with the X-ray crystallographic, cryo-EM and SDM data may enhance our understanding of the active site regions and catalytic mechanism of these viral RdRps and will help to assess not only the effectiveness of the existing drugs but also to design and develop new antiviral drugs for these enzymes.

ACKNOWLEDGMENTS

The author wishes to thank Dr. H. Shakila, Professor& Head, Department of Molecular Microbiology, School of Biotechnology, Madurai Kamaraj University, Madurai, for useful suggestions on the manuscript.

COMPETING INTERESTS

The author has declared that no competing interests exist.

REFERENCES

1. Kenneth KS Ng, Jamie J. Arnold, and Craig E. Cameron. Structure-function relationships among RNA-Dependent RNA polymerases. Curr Top Microbiol Immunol. 2008;320:137–156.
2. Wang Q, Wu J, Wang H, Gao Y, Liu Q, Mu A, Ji W, et al. Structural basis for RNA replication by the SARS-CoV-2 polymerase. Cell. 2020;182:417–428.
3. Alberts B, Johnson J, Lewis J, Raff M, Roberts K, Walter P. Molecular biology of the cell. New York: Garland Science; 2003.
4. Flint SJ, Enquist LW, Racaniello VR, Skalka AM. Principles of virology: molecular biology, pathogenesis, and control of animal viruses. Molecular biology. Washington, DC: ASM Press; 2004.
5. Elena SF, Sanjuan R. Adaptive value of high mutation rates of RNA viruses: Separating causes from consequences. J. Virol. 2005;79:11555–11558.
6. Castro C, Arnold JJ, Cameron CE. Incorporation fidelity of the viral RNA-dependent RNA polymerase: a kinetic, thermodynamic and structural perspective. Virus Res. 2005;107:141–149.
7. Palanivelu P. Analyses of the Spike Proteins of Severe Acute Respiratory Syndrome-Related Coronaviruses. Microbiol Res J Int. 2020; 30:32-50.
8. Chan JFW, Kok KH, Zhu Z, Chu H , To KKW , Yuan S , Yuen KY. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect. 2020;9:221-236.
9. McBride R, Fielding BC. The role of severe acute respiratory syndrome (SARS)-coronavirus accessory proteins in virus pathogenesis. Viruses. 2012;4:2902-2923.
10. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. In Maier HJ, Bickerton E, Britton P, editors. Coronaviruses. Met Mol Biol. 1282. Springer. 2015;1–23.
DOI: 10.1007/978-1-4939-2438-7_1.
11. Thoms M, Buschauer R, Ameismeier M, Koepke L, Denk T, Hirschenberger M. Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. Science. 2020;369:1249–1255.
12. Wolff G, Limpens RWAL, Zevenhoven-Dobbe JC, Laugks U, Zheng S, de Jong AWM, Koning RI, et al. A molecular pore spans the double membrane of the coronavirus replication organelle. Science. 2020;369:1395-1398.
13. Lei J, Kusov Y, Hilgenfeld R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. Antiviral Res. 2018;149:58–74.
14. Sparrer KM, Gack MU. Intracellular detection of viral nucleic acids. Curr. Opin. Microbiol. 2015;26:1-9.
15. Bianchi M, Benvenuto D, Giovanetti M, Angeletti S, Ciccozzi M, Pascarella S. Sars-CoV-2 Envelope and Membrane Proteins: Structural differences Linked to Virus Characteristics? BioMed Res. Int. 2020;2020:1-6.

16. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579:265-269.
17. Hillen HS, Kokic G, Farnung L, Dienemann C, Tegunov D, Cramer P. Structure of replicating SARS-CoV-2 polymerase. bioRxiv preprint; 2020. Available:doi.org/10.1101/2020.04.27.063180.
18. Subissi L, et al. SARS-CoV ORF1b-encoded nonstructural proteins 12-16: replicative enzymes as antiviral targets. Antiviral Res. 2014;101:122-130. DOI: 10.1016/j.antiviral.2013.11.006.
19. Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, Wang T, et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science. 2020;368:779-782.
20. Zhang WF, Stephen P, Thériault JF, Wang R, Lin SX. Novel coronavirus polymerase and nucleotidyl-transferase structures: potential to target new outbreaks. J Phys Chem Lett. 2020;11:4430-4435.
21. Frediansyah A, Tiwari R Sharun K, Dhama K, Harapan H. Antivirals for COVID-19: A critical review. Clinical Epidemiology and Global Health; 2021. doi.org/10.1016/j.cegh.2020.07.006
22. Palanivelu P. DNA polymerases – An insight into their active sites and mechanism of action, In: Recent Advances in Biological Research. Sciencedomain International Book Publishers, UK. 2019;1: 1-39, ISBN: 978-81-934224-4-1, DOI:10.9734/bpi/rabr/v1,
23. Palanivelu P. Single Subunit RNA Polymerases: An Insight into their Active Sites and Catalytic Mechanism, In: Advances and Trends in Biotechnology and Genetics. 2019;1:1-38. Sciencedomain International Book Publishers, UK. ISBN: 978-93-89246-59-9, DOI:10.9734/bpi/atbg/v1,
24. Palanivelu P. Multi-subunit RNA Polymerases of Bacteria - An insight into their active sites and catalytic mechanism. Indian J Sci Technol. 2018;11:1-37.
25. Palanivelu, P. Eukaryotic multi-subunit DNA dependent RNA Polymerases: An insight into their active sites and catalytic mechanism. In: Emerging Trends and Research in Biological Science. Sciencedomain International Book Publishers, UK. 2020;1:1-66. ISBN: 978-93-89562-56-9, DOI: 10.9734/bpi/etrbs/v1,
26. Palanivelu, P. Active sites of the multi-subunit RNA polymerases of eubacteria and chloroplasts are similar in structure and function: Recent perspectives. In: Current Research Trends in Biological Science. Sciencedomain International Book Publishers, UK. 2020;2:26-61. ISBN: 978-93-90149-66-7, DOI: 10.9734/bpi/crtbs/v2,
27. Hansen JL, Long AM, Schultz SC. Structure of the RNA-dependent RNA polymerase of poliovirus. Structure. 1997; 5:1109-1127.
28. Sangita V, Prasad BVLS, Selvarajan R. RNA dependent RNA polymerases: Insights from structure, function and evolution. Viruses. 2018;10:76-99.
29. Sankar S, Porter AG. Point mutations which drastically affect the polymerization activity of encephalomyocarditis virus RNA-dependent RNA polymerase correspond to the active site of *Escherichia coli*. DNA polymerase I. J Biol Chem. 1992; 267:10168–10176.
30. Jablonski SA, Luo M, Morrow CD. Enzymatic activity of poliovirus RNA polymerase mutants with single amino acid changes in the conserved YGDD amino acid motif. J Virol. 1991;6:4565-4572.
31. Inokuchi Y, Hirashima A. Interference with viral infection by RNA replicase deleted at the carboxy-terminal region. J. Virol. 1987; 61:3946-3949.
32. Yap TL, Xu T, Chen YL, Malet H, Egloff MP, Canard B, Vasudevan SG, Lescar J. Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-Angstrom resolution. J Virol. 2007;81: 4753–4765.
33. Lehmann KC, Gulyaeva A, Zevenhoven-Dobbe JC, Janssen GMC, Ruben M, Overkleeft HS, van Veelen PA, Samborskiy DV, Kravchenko AA, Leontovich AM, et al.. Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the RNA polymerase-containing protein of all nidoviruses. Nucleic Acids Res. 2015;43:8416–8434.
34. Kostyuk SM, Dragan DL, Lyakhov VO, Rechinsky VL, Tunitskaya BK, Chernov SN, Kochetkov E. Mutants of T7 RNA polymerase that are able to synthesize both RNA and DNA. FEBS Letters. 1995; 369:165-168.

35. Imbert I, Guillemot JC, Bourhis JM, Bussetta C, Coutard B, Egloff MP, Ferron F. et al. A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. EMBO J. 2006;25:4933-4942.
36. te Velthuis AJW, van den Worm SH, Snijder EJ. The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. Nucleic Acids Res. 2012;40:1737-17.
37. Subissia L, Posthumab CC, Colleta A, Dobbeb JCZ, Gorbalenyab AE, Decrolya E, Snijderb EJ, Canarda B, Imberta I, One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. Proc Natl Acad Sci (USA). 2014;E3900–E3909.
38. Schmidt ML, Hoenen T. Characterization of the catalytic center of the Ebola virus L polymerase. PLoS Negl Trop Dis. 2017;11:1-14.
39. Tchesnokov EP, Raeisimakiani P, Ngure M, Marchant D, Götte M. Recombinant RNA-dependent RNA polymerase complex of ebola virus. Sci Rep. 2018;8:3970-3979.
40. Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. Cell. 1999; 98:811-824.
41. Woody AY, Eaton SS, Osumi-Davis PA, Woody RW, Asp537 and Asp812 in bacteriophage T7 RNA polymerase as metal ion-binding sites studied by EPR, flow-dialysis, and transcription. Biochemistry. 1996;35:144-152.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/64238