International Journal of Biochemistry Research & Review

30(3): 33-62, 2021; Article no.IJBCRR.70674
ISSN: 2231-086X, NLM ID: 101654445

An Overview of the Proofreading Functions in Bacteria and in Severe Acute Respiratory Syndrome-Coronaviruses

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, analyzed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/IJBCRR/2021/v30i330258
Editor(s):
(1) Prof. Cheorl-Ho Kim, Sungkyunkwan University, South Korea.
(2) Dr. Chunying Li, Georgia State University, USA.
(3) Dr. Muhammad Farhan Jahangir Chughtai, NUR International University, Pakistan.
Reviewers:
(1) Lea Spindler, Univerza V, Mariboru Rektorat, Slovenia.
(2) James Mageto, Kenyatta University, Kenya.
(3) Signe Christensen, Lund University, Sweden.
(4) Talha Bin Emran, BGC Trust University, Bangladesh.
Complete Peer review History: https://www.sdiarticle4.com/review-history/70674

Original Research Article

Received 25 May 2021
Accepted 30 July 2021
Published 06 August 2021

ABSTRACT

Aim: To understand the structure-function relationship of the proofreading (PR) functions in eubacteria and viruses with special reference to Severe Acute Respiratory Syndrome-Coronaviruses (SARS-CoVs) and propose a plausible mechanism of action for PR exonucleases of SARS-CoVs.

Study Design: Bioinformatics, biochemical, site-directed mutagenesis (SDM), X-ray crystallographic data were used to study the structure-function relationships of the PR exonucleases from bacteria and CoVs.

Methodology: The protein sequences of the PR exonucleases of various DNA polymerases, and RNA polymerases of SARS, SARS-related and human CoVs (HCoVs) were obtained from PUBMED and SWISS-PROT databases. The advanced version of Clustal Omega was used for protein sequence analysis. Along with the conserved motifs identified by the bioinformatics analysis, the data already available by biochemical, SDM experiments and X-ray crystallographic analysis on these enzymes were used to arrive at the possible active amino acids in the PR exonucleases of these crucial enzymes.

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

Results: A complete analysis of the active sites of the PR exonucleases from various bacteria and CoVs were done. The multiple sequence alignment (MSA) analysis showed many conserved amino acids, small and large peptide regions among them. Based on the conserved motifs, the PR exonucleases are found to fit broadly into two superfamilies, viz. DEDD and polymerase-histidinol phosphatase (PHP) superfamilies. The bacterial DNA polymerases I and II, RNase D, RNase T and ε-subunit of DNA polymerases III belong to the DEDD superfamily. The PR enzymes from SARS, SARS-related CoVs and other HCoVs also essentially belong to the DEDD superfamily. The DEDD superfamily either uses an invariant Tyr or a His as proton acceptor during catalysis. Depending on the proton acceptor, they are further classified into DEDHD and DEDYD subfamilies. RNase T, ε-subunit of DNA polymerases III and the SARS, SARS-related CoVs and other HCoVs belong to DEDHD subfamily. However, the SARS, SARS-related CoVs and other HCoVs showed additional zinc finger motifs (ZFMs) in their active sites. DNA polymerases I, II and RNase D belong to DEDYD subfamily. The bacterial DNA polymerases X, YcdX phosphoesterases and the co-editing exonuclease of DNA polymerases III belong to the PHP superfamily. Based on the MSA, X-ray crystallographic analyses and SDM experiments, the proposed active-site proton acceptor is Tyr/His in DEDDY/H subfamilies and His in PHP superfamily of PR exonucleases.

Conclusions: Based on the similarities of active site amino acids/motifs, it may be concluded that the DEDD and PHP superfamilies of PR exonucleases should have evolved from a common ancestor but diverged very long ago. The biochemical properties of these enzymes, including the four conserved acidic amino acid residues in the catalytic core, suggest that the CoVs might have acquired the exonuclease function, possibly from a prokaryote. However, the presence of two zinc fingers in the PR active site of the SARS, SARS-related CoVs and other HCoVs sets their PR exonucleases apart from other homologues.

Keywords: Proofreading exonucleases; coronaviruses; SARS-CoVs; DNA polymerases; RNase D, RNase T; DNA polymerase X; ExoN active site; ExoN catalytic Mechanism.

1. INTRODUCTION

Maintenance of genome stability is very important for all living organisms and relies mainly on the DNA and RNA polymerases which replicate the genomes. They replicate the genomes faithfully and thus, preserve and maintain the blueprint of life in all living cells. An in-depth analysis of these crucial catalysts of life, not only reveal fundamental information about their emergence, but also on the evolution of life on earth. Interestingly, not only the living cells but also the non-living entities like DNA and RNA viruses also possess these important enzymes. The DNA and RNA polymerases exhibit strong discrimination for NTPs and dNTPs and rarely insert a wrong nucleotide during replication of the genomes and hence the error rate in DNA or RNA synthesis is very, very minimum and is usually in the order of ~10^{-6}-10^{-9} and~10^{-4} to 10^{-6}, respectively. Even one mistake in critical areas is detrimental to the survival of organisms. Therefore, these crucial enzymes are invariably associated with a PR mechanism to correct any insertion error(s) during genome replication. These PR exonucleases belong to 3'→5' types, and they excise any wrongly added nucleotide from the 3'-growing end, and thus, helping the polymerases to perform error-free genome replication. When a mismatch is encountered by the DNA or RNA polymerases during replication, the polymerases stall/pause, which in-turn activates the PR function which promptly excises the mismatch. Following the excision of the wrong base, the correct base is inserted and replication proceeds. This important PR step in living organisms ensures the original DNA/RNA template is copied without any mistake and passed on to the next generation. These PR enzymes are located either as a part of the replicase on the same polypeptide as a multifunctional enzyme (MFE) or as an independent subunit of a multienzyme complex (MEC). For example, in bacterial DNA polymerases I, three different enzymes are found on a single polypeptide as three distinct domains and exhibit three different activities, viz. i) polymerization, ii) proofreading and iii) DNA repair. The second type of PR exonucleases exists as an independent subunit of a multienzyme complex (MEC), e.g., ε-subunit of the bacterial DNA polymerases III (also known as replicases) [1 and references therein, 2]. To have a holistic view on these important PR enzymes in biological systems, including the one from the SARS, SARS-related and HCoVs an overview of these enzymes is presented in this communication.

1.1 PR Functions in Biological Systems

The PR exonucleases are an important class of exonucleases. They are ubiquitous in biological systems and are reported from viruses, bacteria, fungi, plants, animals, etc. The PR function is not only associated with nucleic acid polymerases like DNA and RNA polymerases, but also associated with other nucleic acid modifying enzymes. Based on the active site amino acids, they are broadly classified into two superfamilies, viz. DEDD and PHP [2,3] Most of the bacterial and CoV DNA/RNA polymerases-associated PR exonucleases use four acidic amino acids, DEDD, for metal-binding and catalysis and hence belong to the DEDD superfamily whereas the DNA polymerases X, DNA polymerase III co-editing exonuclease [4] and YcdX phosphoesterases [5] use essentially His residues for metal-binding and catalysis. In this communication, the PR exonucleases belonging to the two different superfamilies are discussed in detail.

1.1.1 PR Exonucleases of DEDD and PHP superfamilies

The DEDD superfamily consists of two subfamilies, viz. DEDDy and DEDDh, depending upon whether they employ an invariant Y or a H as the proton acceptor during catalysis [2]. At least three different DNA polymerases involve in DNA repair and replication processes in prokaryotes. They are DNA polymerase I (encoded by *polA),* DNA polymerase II (encoded by *polB*) and DNA polymerase III, (a MEC). The DNA polymerases I, II and RNase D, belong to the DEDDY subfamily whereas the proofreading ε-subunit of the DNA polymerase III, RNase T, Exons of the RNA-dependent RNA polymerases (RdRps) of SARS, SARS-related CoVs and other HCoVs belong to the DEDDH subfamily. These two subfamilies and the PHP exonuclease superfamily are analyzed and discussed in detail.

2. MATERILAS AND METHODS

The protein sequences of the PR exonucleases of various DNA polymerases, RNases D, RNases T and RNA polymerases of SARS, SARS-related and human CoVs (HCoVs) were obtained from PUBMED and SWISS-PROT databases. The advanced version of Clustal Omega was used for protein sequence analysis. Along with the conserved motifs identified by the bioinformatics analysis, the data already available by biochemical, SDM experiments and X-ray crystallographic analysis on these enzymes were used to arrive at the possible active amino acids in the PR exonucleases.

3. RESULTS AND DISCUSSION

The PR function in eubacterial DNA polymerases I exists as an independent domain on the same polypeptide. The DNA polymerase I of *E. coli* is studied in great detail [1 and references therein]. It is a MFE and consists of three enzymes viz. i) 5'-3' exonuclease (DNA repair function), ii) 3'-5' exonuclease (PR function) and iii) DNA polymerase and are located in three independent domains of the same polypeptide (Fig. 1). The last two domains are also known as Klenow polymerase, and the distance between them is found to be ~30 Å.

Based on the conserved active site amino acids, this PR exonuclease is classified into DEDDy type [2]. Fig. 2 shows the MSA of the DNA polymerases I from different bacteria. Only the PR domains, from amino acids from 324-517, (numberings from the *E. coli* DNA polymerases I) and the polymerase region are shown here. (The *E. coli* enzyme is highlighted in yellow and the possible metal-binding sites are highlighted in green). The MSA analysis shows that the PR domain is almost completely conserved in all bacteria, except for few minor variations. At least four metal-binding sites are observed, and all the four are found in the completely conserved blocks. Two of the metal-binding sites, viz. –DTE- (355-357) and –DAD- (501-503), were proved to be essential for PR exonuclease activity by SDM and X-ray crystallographic analyses in the *E. coli* DNA polymerases I (marked in red) [1 and references therein]. Interestingly, the –**D**T**E**- site is found to be a fusion site where two –**D**X**D**- types of motifs are fused as one single site -F**D**T**E**T**D**S- (where the D is replaced by an equivalent amino acid E) and both the motifs are followed by a hydroxyl amino acid, T or S and the two Ts lie in-between. The typical -DEDD- with 4 invariant acidic amino acids is found in all the 3'-5' exonucleases from bacterial DNA polymerases I and highlighted. The invariant –YA- template-binding pair (highlighted in yellow) suggests that they are all strictly template-dependent enzymes. The PR exonuclease with the four invariant acidic amino acids with an identifiable pattern -DxE→D→Y→D- belongs to the DEDD exonuclease superfamily and to the dnaQ-Y subfamily (Fig. 2).

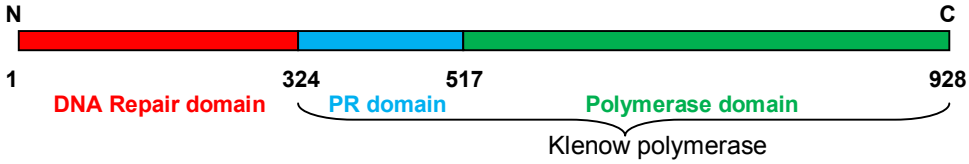


Fig. 1. Dissection of the 3 functional domains of *E. coli* DNA polymerase I

Clustal (O) of PR 3'-5' Exonucleases of bacterial DNA polymerases I (only the PR exonuclease and polymerase regions are shown).

```
// End of the polymerase I
tr|A0A514EYQ8|A0A514EYQ8_9ENTR      HELMENSTTLAVPLLVEVGSGENWDQAH     928
tr|A0A1V0LLD5|A0A1V0LLD5_9ENTR      HELMESSTTLDVPLLVEVGSGQNWDQAH     928
tr|A0A1C4GJM7|A0A1C4GJM7_9ENTR      HELMENSTTLDVPLLVEVGSGQNWDQAH     928
tr|F0JW51|F0JW51_ESCFE              HQLMENCTRLDVPLLVEVGSGENWDQAH     928
tr|V0Y9R8|V0Y9R8_ECOLX              HQLMENCTRLDVPLLVEVGSGENWDQAH     928
tr|I6DHG2|I6DHG2_SHIBO              HQLMENCTRLDVPLLVEVGSGENWDQAH     928
sp|P00582|DPO1_ECOLI                HQLMENCTRLDVPLLVEVGSGENWDQAH     928
tr|A0A3P6LP28|A0A3P6LP28_SHIDY      HQLMENCTRLDVPLLVEVGSGENWDQAH     928
sp|Q9F173|DPO1_SALTY                HQLMENCTRIDVPLLVEVGSGENWDQAH     928
tr|A0A2I8SCZ7|A0A2I8SCZ7_9ENTR      HQLMENETQIDVPLLVEVGSGENWDQAH     928
                                    *:***. * : **********:******
```

Fig. 2. MSA of the PR 3'-5' Exonucleases of bacterial DNA polymerases I

A0A514EYQ8_9ENTR, Raoultella electrica
A0A1C4GJM7_9ENTR, Kosakonia oryziphila
V0Y9R8_ECOLX, Escherichia coli
P00582|DPO1_ECOLI, Escherichia coli (strain K12)
Q9F173|DPO1_SALTY, Salmonella typhimurium
A0A1V0LLD5_9ENTR, Kosakonia radicincitans
F0JW51_ESCFE, Escherichia fergusonii ECD227
I6DHG2_SHIBO, Shigella boydii
A0A3P6LP28_SHIDY, Shigella dysenteriae
A0A2I8SCZ7_9ENTR, Citrobacter freundii

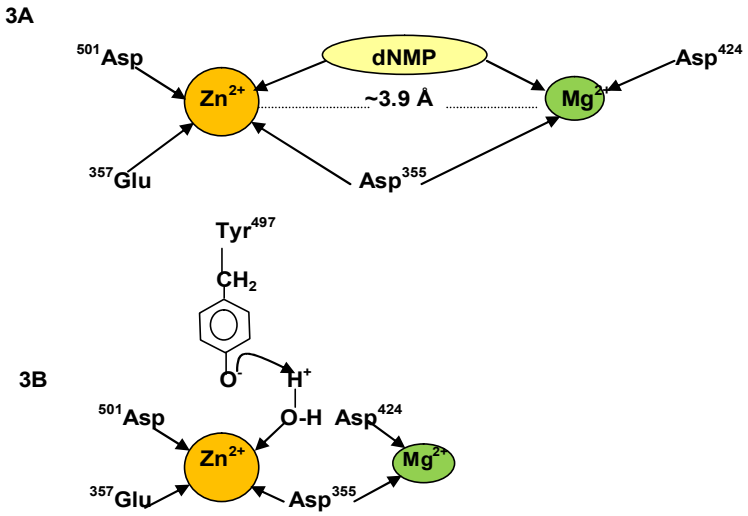


Fig. 3. A Schematic diagram showing the subsites-A and -B of the PR exonuclease of E. coli DNA polymerase I. (Active sites amino acids are placed based on the crystallographic and SDM data)
Fig. 3B shows only the proposed amino acids at the exonuclease active site with a water molecule as the 4th ligand. (All the four acidic amino acids of the DEDD superfamily are shown here)

3.1 Active Site Analyses of the PR Exonuclease of DNA polymerase I

The PR exonuclease activity of the DNA polymerase I of *E. coli* is one of the most well-studied enzymes in this class [6. 7]. The active site of the PR exonuclease of the *E. coli* DNA polymerase I was analysed by genetics and SDM experiments and also by crystallographic studies. It was found that the PR exonuclease active site (EAS) essentially consisted of two sites, viz. a dNMP site and a metal-binding site. Therefore, dNMPs could inhibit the exonuclease reaction by product inhibition. The metal-binding site consisted of two subsites, viz., subsite-A and subsite-B and thus, EAS can bind two divalent metal ions. The presence of two divalent metal ion binding sites was further confirmed by anomalous scattering difference Fourier analysis of the wild-type enzyme with the ligands [6].

The subsite-A is coordinated by three amino acids, viz., Asp^{355}, Glu^{357} and Asp^{501} and the dNMP-phosphate provides the fourth ligand. Usually a Zn^{2+} is associated to the subsite-A. The second metal-binding site, subsite-B, is mainly coordinated by Asp^{424} and to the divalent metal ion Mg^{2+}. The subsite-B is located between

dNMP-phosphate and the carboxylate of Asp^{424} (Fig. 3A). The Zn^{2+} binding subsite-A was found to be very close to the 3' O-of the susceptible bond to be cleaved, and the Mg^{2+} binding subsite-B is very close to subsite-A. X-ray crystallographic data showed that the distance between the two metal atoms is ~ 3.9 Å in *E. coli* PR exonuclease (Fig. 3A) [6].

Further insights into the amino acids that constitute the EAS, were provided by SDM experiments by Joyce and Steitz [8].

a) In a double mutant with $Asp^{355}\rightarrow$Ala and $Glu^{357}\rightarrow$Ala, both the dNMP binding site and the metal-binding site A (Zn^{2+}) were completely abolished. This mutant protein had lost the exonuclease activity, but exhibited the polymerase activity. This suggested that the dNMP site is coordinating by both the metal-binding sites (Fig. 3A) [8].
b) In the second SDM experiment, the Asp^{424} was replaced by Ala ($Asp^{424}\rightarrow$Ala). In this mutant enzyme, the metal-binding site B (Mg^{2+}) was abolished and exhibited no exonuclease activity (the mutant protein, $D^{424}\rightarrow$A, did not bind to the metal ion in subsite B). However, in this mutant enzyme also the polymerase activity was found to be preserved [8].

These data suggest that the metal ions play a direct role during PR activity. The SDM studies have further shown that both the metal-binding sites are functionally connected and in the absence of one, the other cannot function. The Zn^{2+} binding site possibly involves in catalysis and the Mg-binding site, bind to dNMP-phosphate and link the dNMP site. The Zn^{2+} ligands, viz. D^{355}, E^{357}, D^{501} with a water molecule were found within ~ 2.0 Å distances [8]. Furtehrmore, substitution of Asp^{355} and Glu^{357} in the *E. coli* polymerase I yielded an enzyme devoid (<0.01% remaining) of exonuclease activity, while retaining its overall structure [6]. The Tyr^{497} is placed as the proton acceptor as it is a completely conserved in the highly conserved block and is in the equivalent position to the His^{162} of the PR exonuclease of the DNA polymerase III ε-subunit (dnaQ-H family). Furthermore, the PR exonuclease of DNA polymerase I belongs to dnaQ-Y family where an invariant Y is proposed to involve in deprotonation of water molecule similar to the H^{162}, an invariant amino acid in the dnaQ-H family performing the same function.

A nucleophilic attack on the phosphorous atom of the terminal nucleotide is postulated to be carried out by a hydroxide ion that is activated by one divalent metal, while the expected pentacoordinate transition state and the leaving oxyanion are stabilized by a second divalent metal ion that is placed 3.9 Å away from the first metal ion [9]. Of particular importance is the mutant protein $D^{424}\rightarrow$A, which showed no measurable exonuclease activity. Not only this implies an important catalytic role for the metal ion B, but it also allows the preparation of a stable complex with a single-stranded DNA substrate. Therefore, it was concluded that the chemical catalysis of the hydrolytic phosphoryl transfer reaction is promoted by the two metal ions and a water molecule which is coordinated to the Zn^{2+} [6].

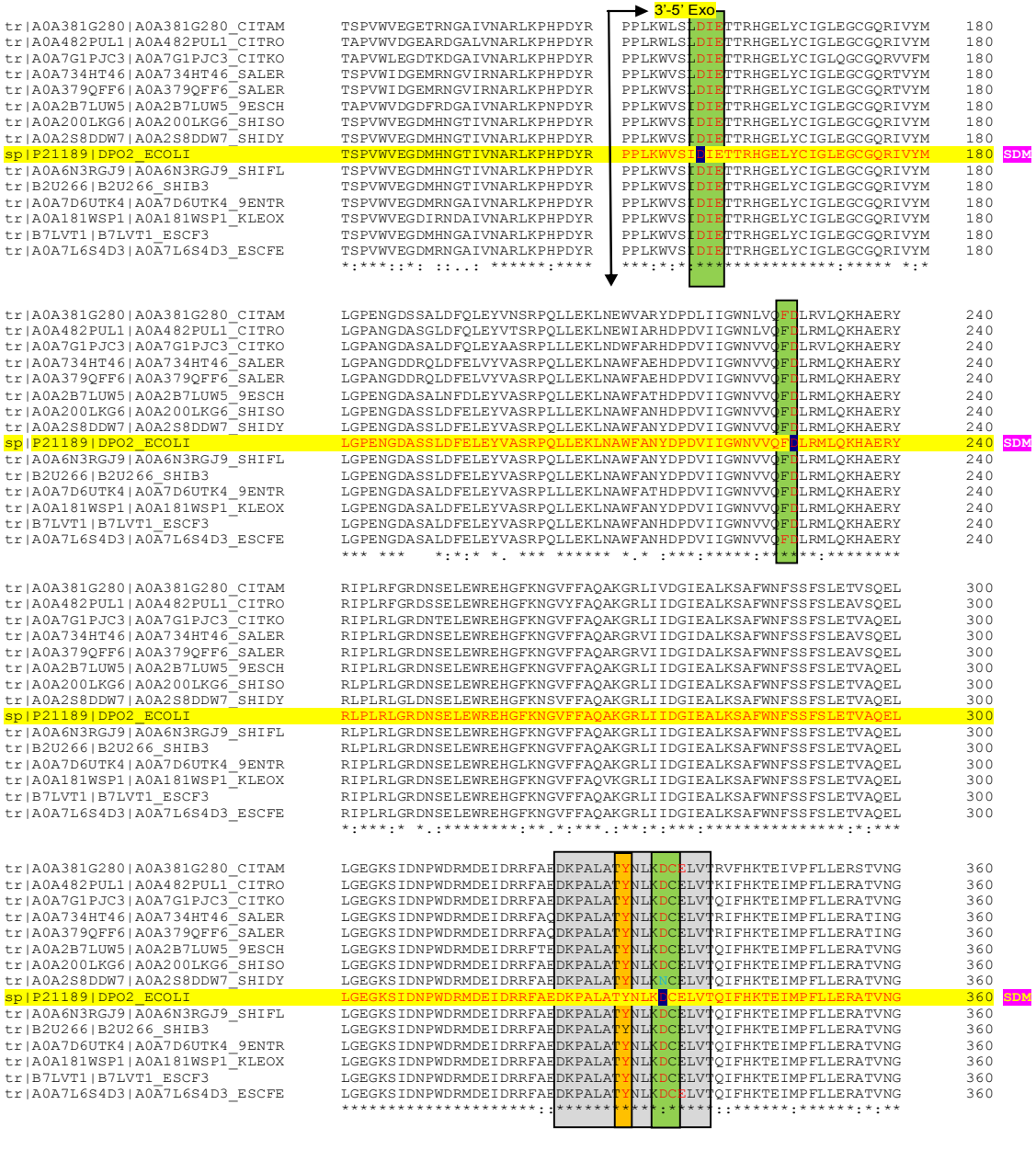
Furthermore, the pH dependence of the 3'-5' exonuclease reaction is consistent with a mechanism in which nucleophilic attack on the terminal phosphodiester bond is initiated by a hydroxide ion coordinated to one of the enzyme-bound metal ions [7]. The properties of the mutant proteins suggest that one metal ion plays a role in substrate binding while the other is involved in catalysis [6]. It is interesting to note that a complete absence of any C in this PR domain suggests that the Zn^{2+} is not coordinated by Cs as reported in other PR exonucleases discussed elsewhere. Based on the crystallographic data and SDM analysis, the proposed amino acids at the PR exonuclease active site is shown in Fig. 3B

3.2 PR Function in DNA Polymerases II (DEDYD)

The second enzyme that shows an intrinsic PR exonuclease function is the bacterial DNA polymerases II. The DNA polymerase II from *E. coli* is one of the most well-studied enzyme among this class [1 and references therein, 7]. It is encoded by *polB* gene and consists of 783 amino acids with a molecular mass of ~90 kDa. It is a member of the Family B DNA polymerases, or otherwise known as repair polymerases. Like DNA polymerase I, it exists as a monomer and the catalytic core consists of the typical structural domains, viz. palm, fingers, and thumb as reported in other DNA/RNA polymerases. The enzyme exhibits both 5′→3′ DNA synthesis and 3′→5′ PR exonuclease activity. The DNA polymerases II can extend primers in a variety of lesions, which is known as translesion synthesis. Wang and Yang [10] have shown that the amino

acids 147 to 367 comprise the 3'-5' PR exonuclease domain (highlighted in red) and 368 to 783 involve in polymerase function in *E. coli* polymerase II. As the D^{335}→N mutant lost its exonuclease activity, it is implicated in the catalysis. The MSA analysis shows that the exonuclease domain is highly conserved in all bacterial DNA polymerases II (The *E. coli* sequence is highlighted in yellow). The PR exonuclease contains the four invariant acidic amino acids with an identifiable pattern -DxE→D→Y→D- and hence belongs to the DEDD exonuclease superfamily and to dnaQ-Y subfamily (Fig. 4).

CLUSTAL O (1.2.4) MSA of DNA Polymerases II (only the PR exonuclease and polymerase active site regions are shown)



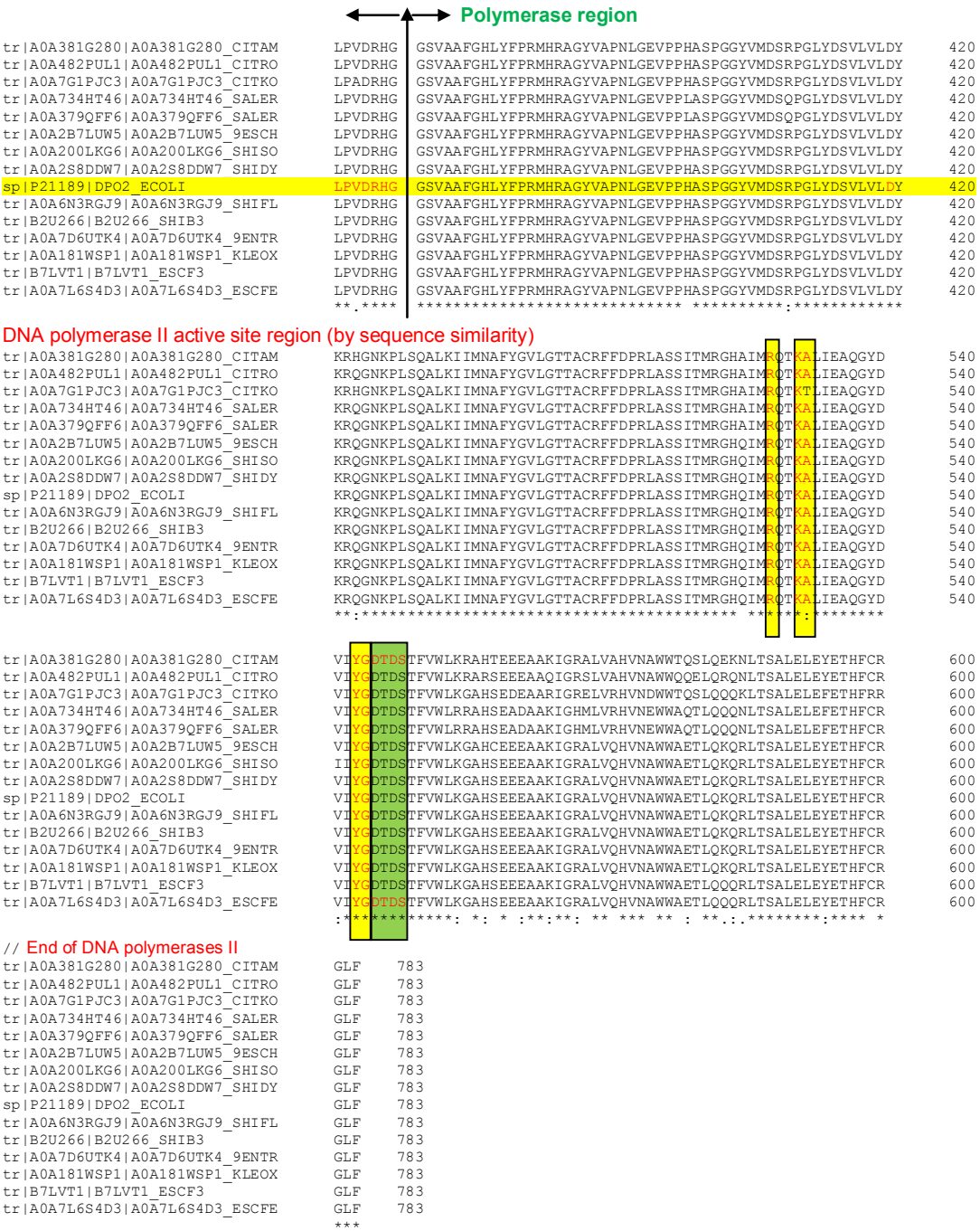


Fig. 4. MSA of the PR Exonuclease domain of bacterial DNA polymerases II

A0A381G280_CITAMCitrobacter amalonaticus
A0A482PUL1_CITRO, Citrobacter rodentium
A0A7G1PJC3_CITKO, Citrobacter koseri
A0A734HT46_SALER, Salmonella enterica subsp. Salamae
A0A379QFF6_SALER, Salmonella enterica
A0A2B7LUW5_9ESCH, Escherichia marmotae
A0A200LKG6_SHISO, Shigella sonnei
A0A2S8DDW7_SHIDY, Shigella dysenteriae
P21189|DPO2_ECOLI, Escherichia coli
A0A6N3RGJ9_SHIFL, Shigella flexneri
B2U266_SHIB3, Shigella boydii
A0A7D6UTK4_9ENTR, Enterobacter hormaechei
A0A181WSP1_KLEOX, Klebsiella oxytoca
B7LVT1_ESCF3, Escherichia fergusonii
A0A7L6S4D3_ESCFE, Escherichia fergusonii

3. 2.1 Active Site Analyses of the PR Exonuclease of DNA polymerase II

The probable active site amino acids in the bacterial DNA polymerases II are shown in Fig. 5. The exonuclease amino acids are very similar to the *E. coli* DNA polymerase I as both belong to the DEDD superfamily of exonucleases. The active site amino acids are placed based on the crystallographic and SDM data available on the enzyme (D^{156}→N, D^{335}→N and D^{229}→N are the exo$^-$ mutants) [10]. The Tyr^{331} is placed as the proton acceptor as it is a completely conserved amino acid in highly conserved block which is in the equivalent position to the His^{162} of the PR exonuclease of the DNA polymerase ε subunits. Furthermore, this PR function in DNA polymerase II could also belong to the dnaQ-Y family like DNA polymerase I where an invariant Y could involve in deprotonation of water molecule similar to the H^{162}, an invariant amino acid in the dnaQ-H family performing the same function [10]. It is interesting to note that in both the DNA polymerases, (i.e.), I and II, the proton acceptor Tyr exhibits also a distance conservation, (i.e.), 4 amino acids from the last catalytic Asp (Fig. 5).

3.3 PR Function RNase D Exonucleases (DEDYD)

The RNase D (EC 3.1.13.5), one of the seven exoribonucleases, which involve in the 3'-maturation of several stable RNAs like tRNA, 5S rRNA, and other small structured RNAs, also belong to the DEDD family [2]. They initiate hydrolysis at the 3'-terminus of an RNA molecule and releases 5'-mononucleotides. MSA of the RNase D from different bacteria is shown in Fig. 6. It is not highly conserved like the DNA polymerases I and II (Figs. 2, 4). However, the active site regions are completely conserved. The metal-binding regions are highlighted in light green and the proton acceptor is highlighted in orange. This class of enzyme also shows the four conserved acidic amino acids, viz. DEDD with the Y as the proton acceptor and hence belong to the DEDD superfamily of exonuclease. The pattern found –DxE→D→Y→D- is very similar to the other exonucleases of DEDYD subfamily of enzymes like DNA polymerases I and II.

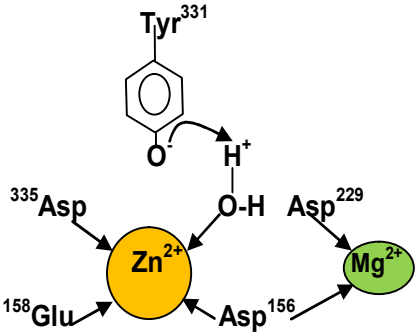


Fig. 5 Proposed amino acids at the active site of the PR Exonuclease of DNA polymerases II of *E. coli*

CLUSTAL O (1.2.4) multiple sequence alignment RNase D (Only the exonuclease domain is shown)

```
sp|A6V8R6|RND_PSEA7               MFVTAPEIQWIRDDASLAQQCREWRTQPYLALDTEFMRVDTFYPAAGLVQVGDGRREWLI       60
tr|A0A6I7I490|A0A6I7I490_KLEPN    --MTAPEIQWIRDDASLAQQCREWRTQPYLALDTEFMRVDTFYPAAGLVQVGDGRQEWLI       58
tr|B7LPP7|B7LPP7_ESCF3            -----MNYQLITTDDALASLCEAVRAFPAIALDTEFVRTRTYYPQLGLIQLFDGEHLALI       55
tr|A0A7H9K3L2|A0A7H9K3L2_9ESCH    ---------MITTDDALASLCEAVRAFPAIALDTEFVRTRTYYPQLGLIQLFDGEHLALI       51
tr|A0A7D6UZ56|A0A7D6UZ56_9ENTR    ---------MITTDDALASLCEAVRAFPAIALDTEFVRTRTYYPQLGLIQLFDGEHLALI       51
tr|A0A659ZP75|A0A659ZP75_SHISO    ---------MITTDDALASLCEAVRAFPAIALDTEFVRTRTYYPQLGLIQLFDGEHLALI       51
sp|P09155|RND_ECOLI               -----MNYQMITTDDALASLCEAVRAFPAIALDTEFVRTRTYYPQLGLIQLFDGEHLALI       55
tr|A0A380AWJ5|A0A380AWJ5_SHIFL    ---------MITTDDALASLCEAVRAFPAIALDTEFVRTRTYYPQLGLIQLFDGEHLALI       51
tr|A0A6I1J9T0|A0A6I1J9T0_9ENTR    ---------MITTDDALASLCEAVRAFPAIALDTEFVRTRTYYPQLGLIQLFDGEHLALI       51
tr|A0A5Q3TIV6|A0A5Q3TIV6_9ENTR    ---------MITTDDALASLCEAVRAFPAIALDTEFVRTRTYYPQLGLIQLFDGEHLALI       51
tr|E7SVN6|E7SVN6_9ENTR            ---------MITTDDALASLCEAVRAFPAIALDTEFVRTRTYYPQLGLIQLFDGEHLALI       51
                                           *  * :**. *.  *: * :*******:*. *:**  **:*: **.:  **
```

```
sp|A6V8R6|RND_PSEA7              DPLLVRDWGPFAELLEDPRVVKVLHACSEDLEVFLRLTGSLPVPLFDTQLAAAYLGMAHS      120
tr|A0A6I7I490|A0A6I7I490_KLEPN   DPLLIQDWSPFAELLEDERVVKVLHACSEDLEVFLRLTGSLPVPLFDTQLAAAYLGMAHS      118
tr|B7LPP7|B7LPP7_ESCF3           DPLTIRDWSPLKSVLRDPAITKFLHAGSEDLEVFLNKFGEMPQPLIDTQVLAAFCGRPMS      115
tr|A0A7H9K3L2|A0A7H9K3L2_9ESCH   DPLGITDWSPLKAILRDPSITKFLHAGSEDLEVFLNVFGELPQPLIDTQILAAFCGRPMS      111
tr|A0A7D6UZ56|A0A7D6UZ56_9ENTR   DPLGISDWSPLKAILRDPSITKFLHAGSEDLEVFLNVFGELPQPLIDTQILAAFCGRPMS      111
tr|A0A659ZP75|A0A659ZP75_SHISO   DPLGITDWSPLKAILRDPSITKFLHAGSEDLEVFLNVFGELPQPLIDTQILAAFCGRPIS      111
sp|P09155|RND_ECOLI              DPLGITDWSPLKAILRDPSITKFLHAGSEDLEVFLNVFGELPQPLIDTQILAAFCGRPMS      115
tr|A0A380AWJ5|A0A380AWJ5_SHIFL   DPLGITDWSPLKAILRDPSITKFLHAGSEDLEVFLNVFGELPQPLIDTQILAAFCGRPMS      111
tr|A0A6I1J9T0|A0A6I1J9T0_9ENTR   DPLGITDWSPLKAILRDPSITKFLHAGSEDLEVFLNVFGELPQPLIDTQILAAFCGRPMS      111
tr|A0A5Q3TIV6|A0A5Q3TIV6_9ENTR   DPLGITDWSPLKAILRDPSITKFLHAGSEDLEVFLNVFGELPQPLIDTQILAAFCGRPMS      111
tr|E7SVN6|E7SVN6_9ENTR           DPLGITDWSPLKSILRDPSITKFLHAGSEDLEVFLNVFGELPQPLIDTQILAAFCGRPMS      111
                                 *** : **.*:  :*.*  :.*.*** ********.  *.:* **:***: **: *   *

sp|A6V8R6|RND_PSEA7              MGYSKLVKEVLDIDLPKDETRSDWLQRPLTEMQMRYAADVQHLAQVYLALDARLSEEKR      180
tr|A0A6I7I490|A0A6I7I490_KLEPN   MGYSKLVKEVLDIDLPKDETRSDWLQRPLTEMQMRYAADVQHLAQVYLALDARLSEEKR      178
tr|B7LPP7|B7LPP7_ESCF3           WGFASMVEEYSGVILDKSESRTDWLARPLTERQCEYAAADVWYLLPITAKLMA--ETEAS      173
tr|A0A7H9K3L2|A0A7H9K3L2_9ESCH   WGFASMVEEYSGVTLDKSESRTDWLARPLTERQCEYAAADVWYLLPIAAKLMV--ETEAS      169
tr|A0A7D6UZ56|A0A7D6UZ56_9ENTR   WGFASMVEEYSGVTLDKSESRTDWLARPLTERQCEYAAADVWYLLPITAKLMV--ETEAS      169
tr|A0A659ZP75|A0A659ZP75_SHISO   WGFASMVEEYSGVTLDKSESRTDWLARPLTERQCEYAAADVWYLLPITAKLMV--ETEAS      169
sp|P09155|RND_ECOLI              WGFASMVEEYSGVTLDKSESRTDWLARPLTERQCEYAAADVWYLLPITAKLMV--ETEAS      173
tr|A0A380AWJ5|A0A380AWJ5_SHIFL   WGFASMVEEYSGVTLDKSESRTDWLARPLTERQCEYAAADVWYLLPITAKLMV--ETEAS      169
tr|A0A6I1J9T0|A0A6I1J9T0_9ENTR   WGFASMVEEYSGVTLDKSESRTDWLARPLTERQCEYAAADVWYLLPITAKLMV--ETEAS      169
tr|A0A5Q3TIV6|A0A5Q3TIV6_9ENTR   WGFASMVEEYSGVTLDKSESRTDWLARPLTERQCEYAAADVWYLLPITAKLMV--ETEAS      169
tr|E7SVN6|E7SVN6_9ENTR           WGFASMVEEYSGVTLDKSESRTDWLARPLTERQCEYAAADVWYLLPITAKLMV--ETEAS      169
                                  *::.:*:*  .: * *.*:*:*** ***** * .*** ** :*  :   * .  . *

// End of RNase D
sp|A6V8R6|RND_PSEA7              SLRGWRRERMGQALLNALESA-- 376
tr|A0A6I7I490|A0A6I7I490_KLEPN   SLRGWRRERMGQALLDALESA-- 374
tr|B7LPP7|B7LPP7_ESCF3           MISGWRGDLMAERLNALLQEYPQ 375
tr|A0A7H9K3L2|A0A7H9K3L2_9ESCH   LISGWRGALMAEALHNLLQEYPQ 371
tr|A0A7D6UZ56|A0A7D6UZ56_9ENTR   LISGWRGELMAEALHNLLQEYPQ 371
tr|A0A659ZP75|A0A659ZP75_SHISO   LISGWRGELMAEALHNLLQEYPQ 371
sp|P09155|RND_ECOLI              LISGWRGELMAEALHNLLQEYPQ 375
tr|A0A380AWJ5|A0A380AWJ5_SHIFL   LISGWRGELMAEALHNLLQEYPQ 371
tr|A0A6I1J9T0|A0A6I1J9T0_9ENTR   LISGWRGELMAEALHNLLQEYPQ 371
tr|A0A5Q3TIV6|A0A5Q3TIV6_9ENTR   LISGWRGELMAEALHNLLQEYPQ 371
tr|E7SVN6|E7SVN6_9ENTR           LISGWRGELMAEALHNLLQEYPQ 371
                                  : ***   *.: *   *:.
```

Fig. 6. MSA of RNase D exonucleases from different bacteria

A6V8R6|RND_PSEA7, Pseudomonas aeruginosa
B7LPP7_ESCF3, Escherichia fergusonii
A0A7D6UZ56_9ENTREnterobacter hormaechei
P09155|RND_ECOLI, Escherichia coli (K12)
A0A6I1J9T0_9ENTR, Enterobacteriaceae bacterium
E7SVN6_9ENTR, Shigella boydii
A0A6I7I490_KLEPN, Klebsiella pneumonia
A0A7H9K3L2_9ESCH, Escherichia marmotae
A0A659ZP75_SHISO, Shigella sonnei
A0A380AWJ5_SHIFL, Shigella flexneri
A0A5Q3TIV6_9ENTR, Salmonella

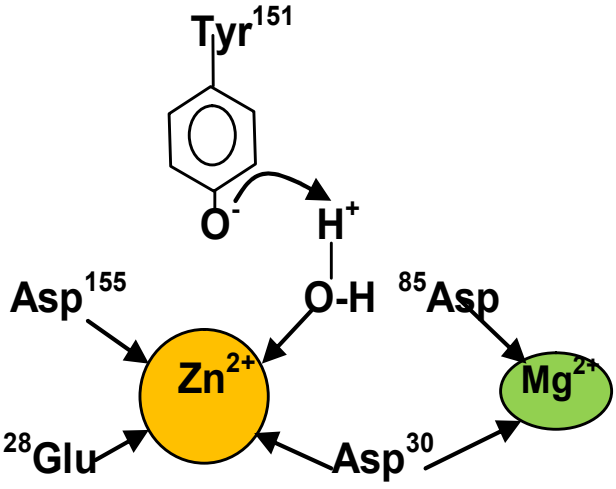


Fig. 7. Proposed amino acids at the active site of the exonuclease of RNase D of *E. coli*

CLUSTAL O (1.2.4) MSA of the ε-subunits of DNA polymerases III

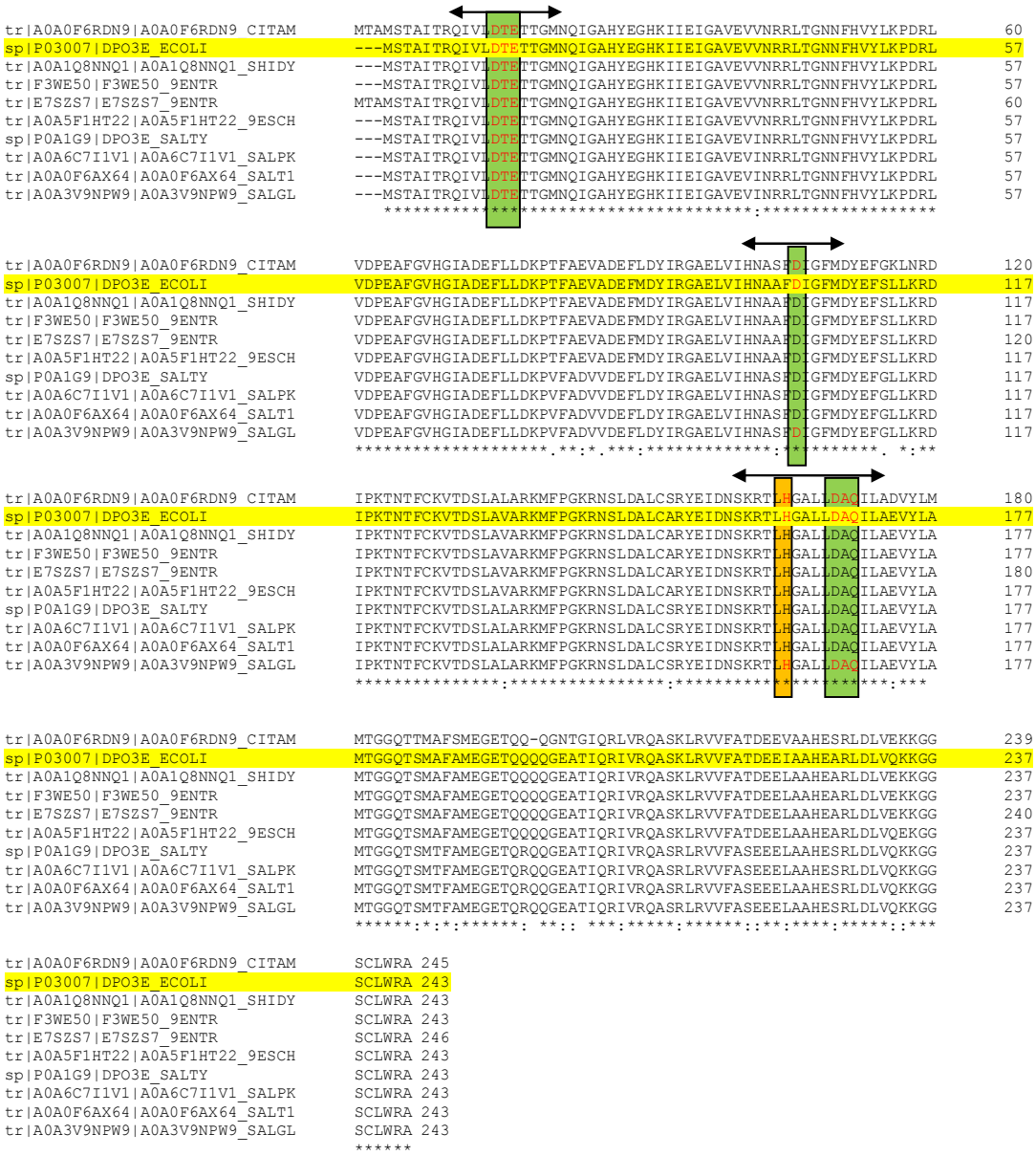


Fig. 8. MSA of PR ε-subunits of DNA polymerases III from different organisms

A0A0F6RDN9_CITAM, Citrobacter amalonaticus
A0A1Q8NNQ1_SHIDY, Shigella dysenteriae
E7SZS7_9ENTR, Shigella boydii
P0A1G9|DPO3E_SALTY, Salmonella typhimurium
A0A0F6AX64_SALT1, Salmonella typhimurium
P03007|DPO3E_ECOLI, Escherichia coli (strain K12)
F3WE50_9ENTR, Shigella boydii
A0A5F1HT22_9ESCH, Escherichia, sp. E4385
A0A6C7I1V1_SALPK, Salmonella paratyphi
A0A3V9NPW9_SALGL, Salmonella gallinarum

The proposed active site amino acids are shown in Fig. 7. Interestingly, all the three DEDDY subfamily of exonucleases maintain a distance conservation between the last D and the proton acceptor Y, and it is only 4 amino acids.

3.4 PR Function in DNA Polymerases III (DEDHD)

Bacterial DNA polymerases III are the third type of polymerases where PR function is reported. The PR function is essential for these types of

enzymes, as they are the replicative enzymes. Unlike the other two polymerases, viz. the polymerases I and II, the polymerases III always exist as MECs with about 10 different subunits. For example, the DNA polymerase III holoenzyme from *E. coli,* the most well characterized in this category, is composed of 10 subunits (α, β, **ε**, θ, δ, δ', γ, τ, χ, ψ), that together with the helicase (DnaB) and the RNA primase (DnaG) form the replisome with a combined molecular weight of ~1 MDa [1 and references therein]. The PR activity is associated with the ε-subunit of the MEC. The ε-subunit which is made up of ~240 amino acid residues, consists of two domains, the N-terminal domain (1-186) with the PR activity and the C-terminal domain bind to the α-subunit of the polymerase. The ε-subunit encoded by *dnaQ* and contains the 3'-5' PR exonuclease catalytic site to edit any misinserted nucleotides by the α-subunit during the synthesis. Fig. 6 shows the MSA of the ε-subunit from different bacteria. The ε-subunits from different bacteria are highly conserved and exhibit over 99% identical residues over the entire sequence. The N-terminal domain contains three conserved exo motifs as marked by arrows. The metal-binding motifs are completely conserved (highlighted in light green) and the possible proton acceptor, H is highlighted in orange. The PR exonuclease contains all of the four invariant acidic amino acids with an identifiable pattern -DxE→D→H→D- and hence belongs to the DEDD exonuclease superfamily and to dnaQ-H subfamily (Fig. 8).

3.4.1 Active Site Analyses of the PR Exonuclease of DNA polymerase III

Fijalkowska and Schaaper [11] found that modification of the two conserved amino acid residues, viz. Asp^{12}→Ala and Glu^{14}→Ala, by SDM experiments resulted in the loss of the exonuclease function and hence suggested to play a role in the coordination of an essential metal ion. Further analysis of the enzyme by Cisneros et al [12] has shown that a water molecule bound to the catalytic metal acts as the nucleophile for the hydrolysis of the phosphate bond. Initially, they observed a direct proton transfer to H^{162}. In a two metal mechanism, the catalytic metal (Me1) is proposed to form an attacking metal-hydroxide which performs a nucleophilic attack on the *α*-phosphate of the nucleotide base to be excised. The second metal (Me2) is termed as the nucleotide binding metal. These observations were further confirmed by X–ray crystallographic analysis of the ε-186 by Hamdan et al [13]. Their results showed that the active site was composed of three residues, D^{12}, E^{14} and D^{167}, and bind to two divalent metals. In addition, H^{162} hydrogen bonds to a water molecule that is coordinated to the catalytic metal (Fig. 9).

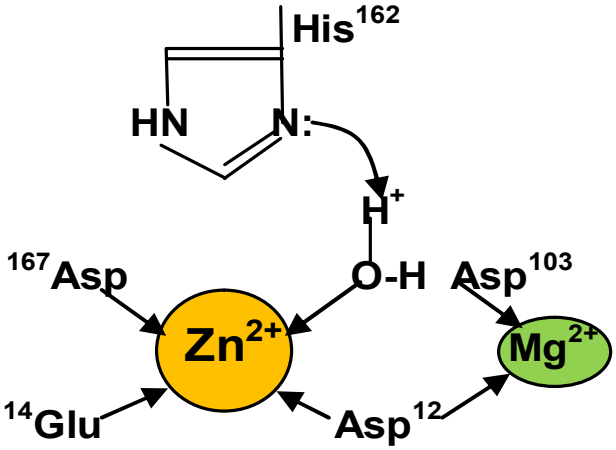


Fig. 9. Proposed amino acids at the active site of the ε-subunits of *E. coli* DNA polymerase III

The ε- exonuclease belongs to the DnaQ-H family with the four active site carboxylates (Asp^{12}, Glu^{14}, Asp^{103}, and Asp^{167}) with the invariant His^{162}. The H^{162} acts as a general base to deprotonate the active site nucleophile. Most important is the substitution of the highly conserved active site tyrosine in enzymes of the DnaQ-Y family (Tyr497 in DNA polymerase I) as the His^{162} in the DnaQ-H family [14].

3.5 PR type 3'-5' Exonuclease Activity in RNases T

Another interesting PR type 3' exonuclease was found not in polymerases, but in the tRNA processing enzyme, RNase T (EC 3.1.13.-) [15]. They are single-strand specific exonucleases which trim short 3' overhangs of a variety of RNA species with extreme sequence specificity, discriminating against cytosine at the 3' end of the substrate and thus leaving one or two nucleotide(s) 3' overhang. They are also responsible for the important end processing reaction in tRNAs, and thus they specifically remove the terminal **AMP** residue from uncharged tRNAs (tRNA-C-C-**A**). Therefore, they play a key role in the maturation of tRNAs. Fig. 8 shows the MSA of the RNase T form different organisms. They are almost completely conserved from N- to C-terminal, with only a few amino acid modifications in the entire sequence. (The *E. coli* enzyme is highlighted in yellow). The metal-binding amino acids are highlighted in light green and the proton acceptor amino acid is highlighted in orange. The MSA analysis shows that they also belong to the DEDD superfamily of 3'-5'exonucleases and are characterized by the presence of four acidic residues, DEDD, making the active site.

CLUSTAL O (1.2.4) MSA of RNase T from various bacterial sources

Fig. 10. MSA of RNase T from different bacteria

A0A564UR04_ESCFE Ribonuclease T, Escherichia fergusonii;
P30014|RNT_ECOLI Ribonuclease T, Escherichia coli (strain K12)
P66683|RNT_SHIFL Ribonuclease T, Shigella flexneri
Q3Z208|RNT_SHISS Ribonuclease T, Shigella sonnei
Q32FB8|RNT_SHIDS Ribonuclease T, Shigella dysenteriae
A0A181WQU7_KLEOX Ribonuclease T, Klebsiella oxytoca
A0A2I8S6X4_9ENTR Ribonuclease T, Citrobacter freundii
A0A482PCL8_CITRO Ribonuclease T, Citrobacter rodentium

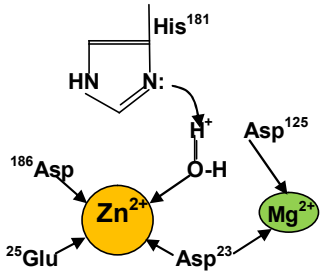


Fig. 11. The proposed amino acids at the active site of RNase T (*E. coli*)

As the exonuclease contains all of the four invariant acidic amino acids with an identifiable pattern -DxE→D→H→D- they are also classified under the DEDD exonuclease superfamily (Fig. 10).

3.5.1 Active Site Analyses of the Exonuclease of RNase T

The *E coli* RNase T has been extensively studied both by X-ray crystallography and SDM experiments [2,15]. Four conserved acidic residues, viz. Asp^{23} and Glu^{25}, Asp^{125} and Asp^{186} were identified in the active site of the enzyme and are also found to be essential for the exonuclease activity. In addition to the above essential amino acid residues, modification of His^{181} by an SDM experiment also abolished the exonuclease activity, suggesting that His^{181} also play an important role in catalysis. Thus, in *E. coli* RNase T, at least five amino acid residues, viz. Asp^{23}, Glu^{25}, Asp^{125}, His^{181} and Asp^{186} are found in the active site (Fig. 11). These residues, together with the substrate, are known to bind two divalent metal ions. The structures of RNase T from *Pseudomonas aeruginosa* and *E. coli* have been solved by Zuo et al [15]. The site A metal ion is coordinated by the three conserved acidic residues, and the site B metal ion is coordinated by the conserved aspartate residue. The B site metal ion in *P. aeruginosa* has an octahedral coordination typical for a magnesium ion whereas the A site metal ion is occupied by a non-magnesium ion (used here a Zn^{2+}). In fact, a water molecule (Water-164) occupying the A site in *P. aeruginosa* RNase T might mimic a non-magnesium metal ion (Zn^{2+}) with 5 potential coordination ligands [15]. Interestingly, all the three DEDDH subfamily of exonucleases maintains a distance conservation between the last D and the proton acceptor H and it is only 5 amino acids. This distance conservation is also maintained in SARS, SARS-related and HCoVS PR exonucleases as well.

4. PR FUNCTION BY PHP SUPERFAMILY OF EXONUCLEASES

A different type of PR function is reported from Family B DNA polymerases. They belong to the PHP superfamily of exonucleases and include DNA polymerases X, DNA polymerases III (co-editing), eukaryotic, YcdX phosphatases, etc. There are at least 4 different X-type of DNA polymerases are also reported from eukaryotes, viz. the terminal transferases and DNA polymerases β, λ, and μ [16, 17]. Unlike the DEDD superfamily, these enzymes use two invariant Hs followed by three acidic amino acids with the general pattern -HxH→E→H→D-. These enzymes which belong to the PHP superfamily are analyzed further for their active site structure(s).

4.1 Bacterial DNA Polymerases X

The DNA polymerases X are ubiquitous like the DNA polymerases I and are reported in wide variety of organisms like viruses, protozoa, archaea, eubacteria, and eukaryotes [3]. They are strictly template-directed DNA polymerases and preferentially act on DNA structures containing gaps from one to a few nucleotides with a phosphate group at the 5' -end of the downstream of DNA fragments. Therefore, they are suggested to participate in the later stages of DNA synthesis like in base excision repair (BER) and in error-prone non-homologous end joining (NHEJ) activity to repair double-stranded breaks [3]. It is interesting to note that these enzymes are structurally different from the DNA polymerases I and II, as the members of this family possess a different type of PR domain, known as the PHP domain. The polymerase domain is found at their N-terminal, whereas the PHP domain is present at the C-terminal end of the polypeptide. The polymerase domains also found to harbour a HNH motif (Fig. 12). Unlike in DNA polymerases I where the PR function is localized in front of the polymerase domain, in X DNA polymerases, it is localized behind the polymerase domain (Figs. 1, 12), i.e., the PR domains are reversed. In addition to the polymerase and PR domains, a HNH endonuclease type of motif (highlighted in red) is also observed towards at the end of the polymerase domain [3].

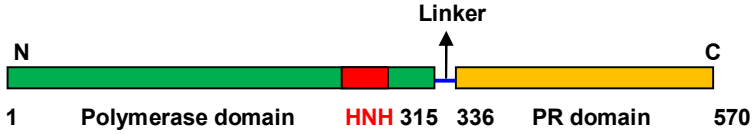


Fig. 12. A schematic diagram of the bacterial DNA Polymerases X

CLUSTAL O (1.2.4) MSA of DNA polymerases X (1-315 Polymerase & 336-570 Exo domains in *B Subtilis*).

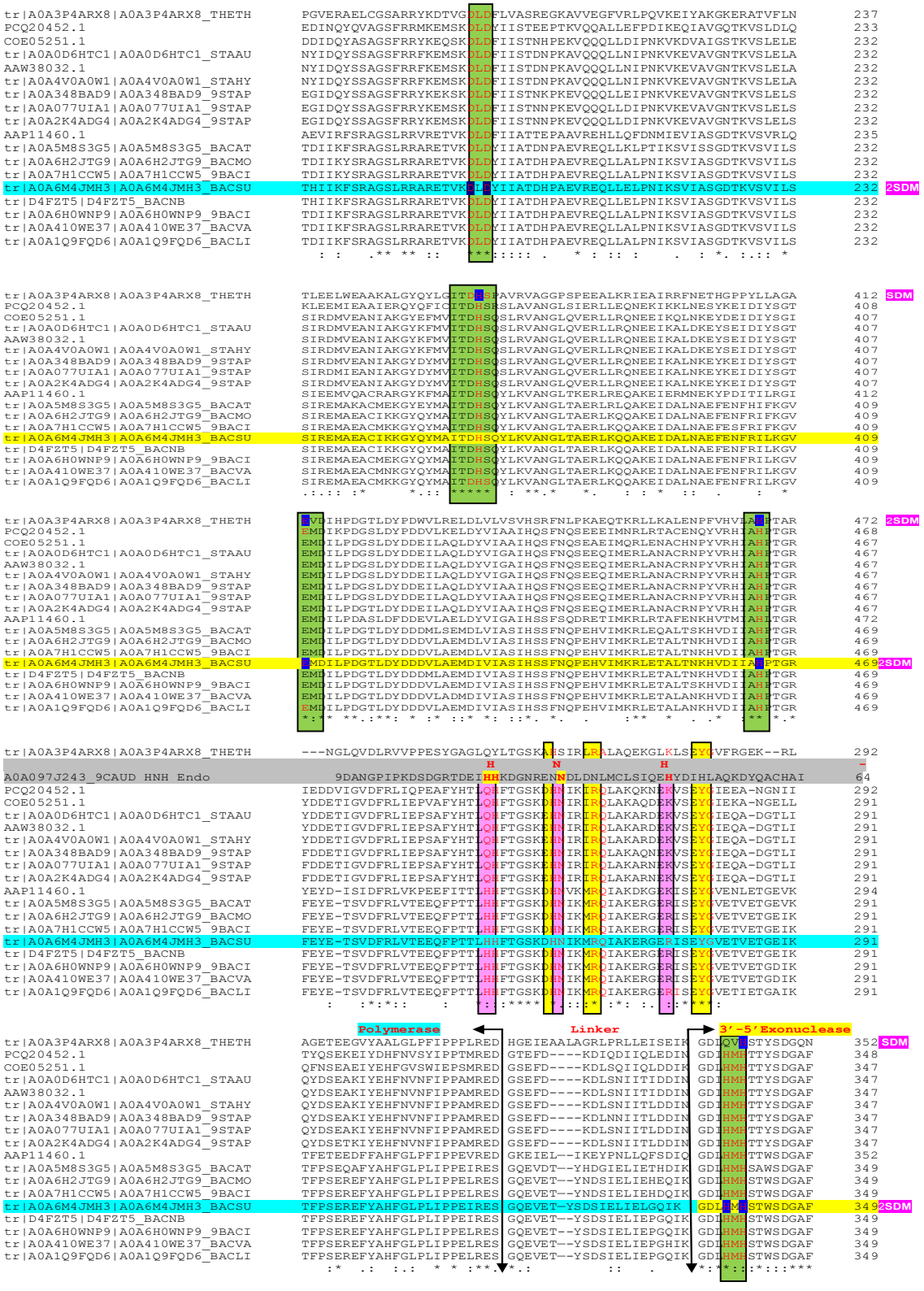


Fig. 13. MSA of the DNA polymerases X from different organisms

A0A3P4ARX8_THETH DNA polymerase beta Thermus thermophilus
PCQ20452.1 DNA polymerase/3'-5' exonuclease PolX [Klebsiella pneumoniae]
COE05251.1 DNA polymerase/3'-5' exonuclease PolX [Staphylococcus warneri]
A0A0D6HTC1_STAAU DNA polymerase beta,Staphylococcus aureus
AAW38032.1 DNA-dependent DNA polymerase family X, Staphylococcus aureus subsp. aureus
A0A4V0A0W1_STAHY DNA polymerase beta,Staphylococcus hyicus
A0A348BAD9_9STAP DNA polymerase beta, Staphylococcus argenteus
A0A077UIA1_9STAP DNA polymerase beta, Staphylococcus schweitzeri
A0A2K4ADG4_9STAP DNA polymerase beta, Staphylococcus schweitzeri
AAP11460.1 DNA polymerase X family, Bacillus cereus
A0A5M8S3G5_BACAT DNA-directed DNA polymerase, Bacillus atrophaeus
A0A6H2JTG9_BACMO DNA-directed DNA polymerase, Bacillus mojavensis
A0A7H1CCW5_9BACI DNA polymerase/3'-5' exonuclease PolX, Bacillus halotolerans
A0A6M4JMH3_BACSU DNA-directed DNA polymerase, Bacillus subtilis
D4FZT5_BACNB DNA-directed DNA polymerase, Bacillus subtilis subsp. Natto
A0A6H0WNP9_9BACI DNA-directed DNA polymerase, Bacillus tequilensis
A0A410WE37_BACVA DNA-directed DNA polymerase, Bacillus vallismortis
A0A1Q9FQD6_BACLI DNA-directed DNA polymerase, Bacillus licheniformis
A0A097J243_9CAUD Homing endonuclease, Enterobacteria phage RB5

4.1.1 Active Site Analyses of PR Exonuclease Domain in Bacterial DNA polymerases X

The PHP domain has been shown to possess the 3′–5′ exonuclease activity and perform the PR function in DNA polymerases X [16-19]. Fig. 13 shows the MSA of the DNA polymerases X form different organisms. Unlike the DNA polymerases I and II, only there are a few conserved motifs among them, suggesting that they are highly diverged during evolution. However, the active site regions are completely conserved (Fig. 13). The X polymerases follow the completely conserved pattern –GSKD **H**$^{-5}$xxxx**RQ**^{1}IAKERGERISE**Y**^{13}**G**V- suggesting the repair is strictly template strand based with the template-binding YG pair with the NTP selecting invariant basic amino acid H at -5 from the catalytic R. (In the DNA polymerase I family, the polymerase catalytic region is slightly different from the X polymerases and is –SEQ **R**$^{-4}$xxx**KA**INFGLI**Y**^{766}**G**I/L/M-). However, the RQ and YG pairs are followed by I//V as in DNA polymerases I. The metal-binding site –DLD- is completely conserved in all X polymerases (highlighted in light green). Within the polymerase domain a HNH homing endonuclease domain is also observed as suggested by Nagpal and Nair [3]. (The active site region of the HNH homing endonuclease domain of Enterobacteria phage RB5 is highlighted in grey). Though the HH---N are completely conserved in all, the last H is not conserved and is replaced with another basic amino acid, K/R in DNA polymerases X, but preceded by an E in all the cases (Fig. 13).

The DNA polymerase domain in X polymerases is located from amino acids 1 to 315 and the 3'-5' exonuclease PR domain is located from amino acids 337-570 (numberings from *Bacillus subtilis* enzyme and highlighted in different colours) (Fig. 13). SDM experiments have shown that the polymerase and PR activities were independent of each other. For example, in a double mutant where both the Ds are replaced by As (^{193}**D**L**D**→**A**L**A**), abolished the polymerase activity [18] whereas in a similar experiment where both the Hs are replaced by As (^{339}**H**M**H**→**A**M**A**), the exonuclease activity was abolished. The metal-dependent exonuclease activity is further confirmed by deletion mutants too. For example, in a deletion mutant 316–570 (mutant ΔPHP), the exonuclease activity was completely abolished [18].

DNA polymerases X of *Thermus thermophilus* was extensively studied by Nakane et al [19]. It possessed both the DNA polymerase (ttPOLXc) and the exonuclease activities belonging to the PHP family, as two independent domains. The enzyme was subjected to SDM experiments by them. The ttPOLXc domain showed Mg^{2+}-dependent DNA polymerase activity but no 3'-5' exonuclease activity. They could hardly detect any 3'–5'exonuclease activity of the mutant enzymes with H^{344}→A, H^{374}→A, H^{468}→A and D^{529}→A mutations. Therefore, they concluded that His^{344} (-QVH^{344}-), His^{374} (-DH^{374}SP-), His^{468} (-AH^{468}P-) and Asp^{529} (-D^{529}AH-) are the most important residues for the 3'–5'exonuclease activity (marked in the margin).

The corresponding amino acids in *B. subtilis* enzyme are placed in the active site (marked in red and highlighted in blue) (Fig. 13). Nakane et al [19] also found that the POLX-core and PHP domains interacted with each other and a mixture of the two domains had Mn^{2+}-dependent 3'–5' exonuclease activity. Importantly, the DNA polymerase exhibited Mg^{2+}-dependent activity and the PHP domain exhibited no exonuclease activity in the presence of Mg^{2+} or Mn^{2+} but exhibited exonuclease activity only with Zn^{2+}, further corroborating the involvement of Zn^{2+} in the PR functions.

The Zn^{2+} atom is coordinated by three invariant His residues, and the fourth ligand is occupied by a water molecule (Fig. 14). Under the polymerase assay conditions, in the presence of Mn^{2+} the Q^{342}→A, D^{349}→A, E^{413}→A and H^{531}→A mutant exhibited stronger 3'-5' exonuclease activity.

The *Bacillus subtilis* DNA polymerases X was also analyzed by SDM experiments by Banos et al. [18]. They found that H^{339} and H^{341} (highlighted in light green) in the HxH motif are shown to be essential for the PR exonuclease active site in the DNA polymerase X. Fig. 14 shows the proposed active site at the PR site in *Bacillus subtilis* DNA polymerases X. The active site amino acids are proposed based on the SDM analysis of *T. thermophilus* and *B. subtilis* X polymerases. They follow the general active site pattern as -HxH→E→H→D-.

4.2 Bacterial YcdX class of Exonucleases

The second group of enzyme is the YcdX types which exhibits a phosphoesterase activity. It is interesting to note, these phosphoesterases do not form a part of polymerases, but also belong to the PHP superfamily because of their structural and sequence similarities [5]. Like the

other PHP enzymes, the YcdX also consists of four conserved sequence motifs that contain invariant histidine and aspartate residues, which are implicated in metal ion coordination and catalysis. X-ray crystallographic data on the *E. coli* YcdX is available [5]. The X-ray crystallographic studies have shown that the catalytic site of YcdX of *E. coli* consists of three Zn atoms and is similar to those enzymes which hydrolyze phosphoester bonds.

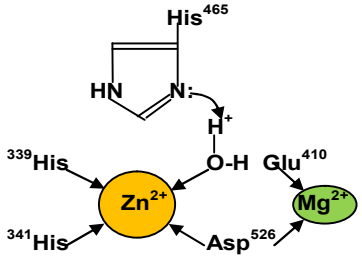


Fig. 14. Proposed amino acids at the active site of DNA polymerases X from *B. Subtilis*

CLUSTAL O (1.2.4) MSA of bacterial YcdX phosphoesterases of PHP superfamily

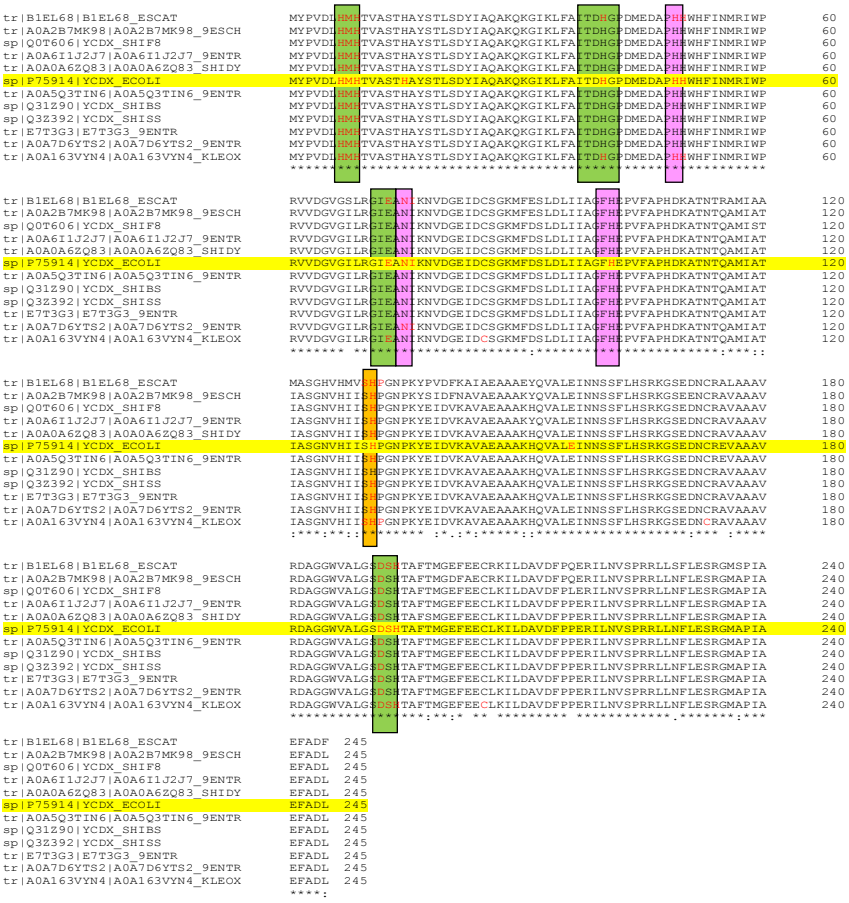


Fig. 15 MSA of YcdX phosphoesterases belonging to PHP superfamily

B1EL68_ESCAT, Escherichia albertii
Q0T606|YCDX_SHIF8, Shigella flexneri
A0A0A6ZQ83_SHIDY, Shigella dysenteria
A0A5Q3TIN6_9ENTR, Salmonella sp.
Q3Z392|YCDX_SHISS, Shigella sonnei
A0A7D6YTS2_9ENTR, Enterobacter hormaechei
A0A2B7MK98_9ESCH, Escherichia marmotae
A0A6I1J2J7_9ENTR, Enterobacteriaceae bacterium
P75914|YCDX_ECOLI, Escherichia coli (strain K12)
Q31Z90|YCDX_SHIBS, Shigella boydii
E7T3G3_9ENTR, Shigella boydii ATCC 9905
A0A163VYN4_KLEOX, Klebsiella oxytoca

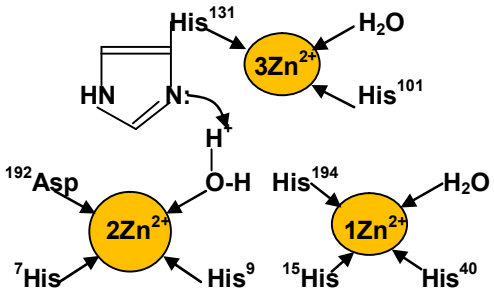


Fig. 16. Proposed amino acids at the active site of *E. coli* YcdX phosphoesterase

They have found that the YcdX had an unusual type of topology with a $\beta_7\alpha_7$barrel type of structure whose C-terminal side had a deep cleft that contained three metal-binding sites which were ligated to the imidazole group of residues His^7, His^9 and His^{15} from motif I; His^{40} from motif II; His^{101} from motif III, His^{131} and His^{194} from motif IV; as well as to the carboxylate group of Glu^{73}, Glu^{156} and Asp^{192}, the latter belonged also to Motif IV. Fig. 15 shows the MSA of the YcdX phosphoesterases which belong to PHP superfamily with the active site pattern as -HxH→E→H→D-. Furthermore, these enzymes are almost completely conserved in their entire sequence only with a few changes. A probable HNH motif is also observed (highlighted in light red) suggesting their possible lateral transfer between genomes of bacterial species.

4.2.1 Active Site Analyses of the YcdX Phosphoesterase

Out of the three Zn atoms, the Zn2 which is coordinated by H^7, H^9, E^{73} and D^{192} and also coordinated to a water molecule could possibly involve in the excision of the nucleotide (Fig. 16). Zn1 is the "high-affinity" site occupied by zinc in the native structure bound to H^{15}, H^{40}, H^{194} and H_20 (marked in red) could likely play the structural role. The proposed active site is based on the X-ray crystallographic data. The catalytic mechanism of YcdX may proceed through the nucleophilic attack of the susceptible phosphorus atom by the water molecule bridging Zn2 and Zn3, which is presumably a hydroxide ion [5 and references therein]. Interestingly, no Mg^{2+} or other divalent metal ions are found in the X-ray crystallographic data. The conserved amino acids, viz. H^{131} of the motif –SH^{131}P- and E^{156} of the motif -E^{156}IN- are based on the sequence similarity to DNA polymerases X.

4.3. Intrinsic PR Activity in the Replicative DNA Polymerases III

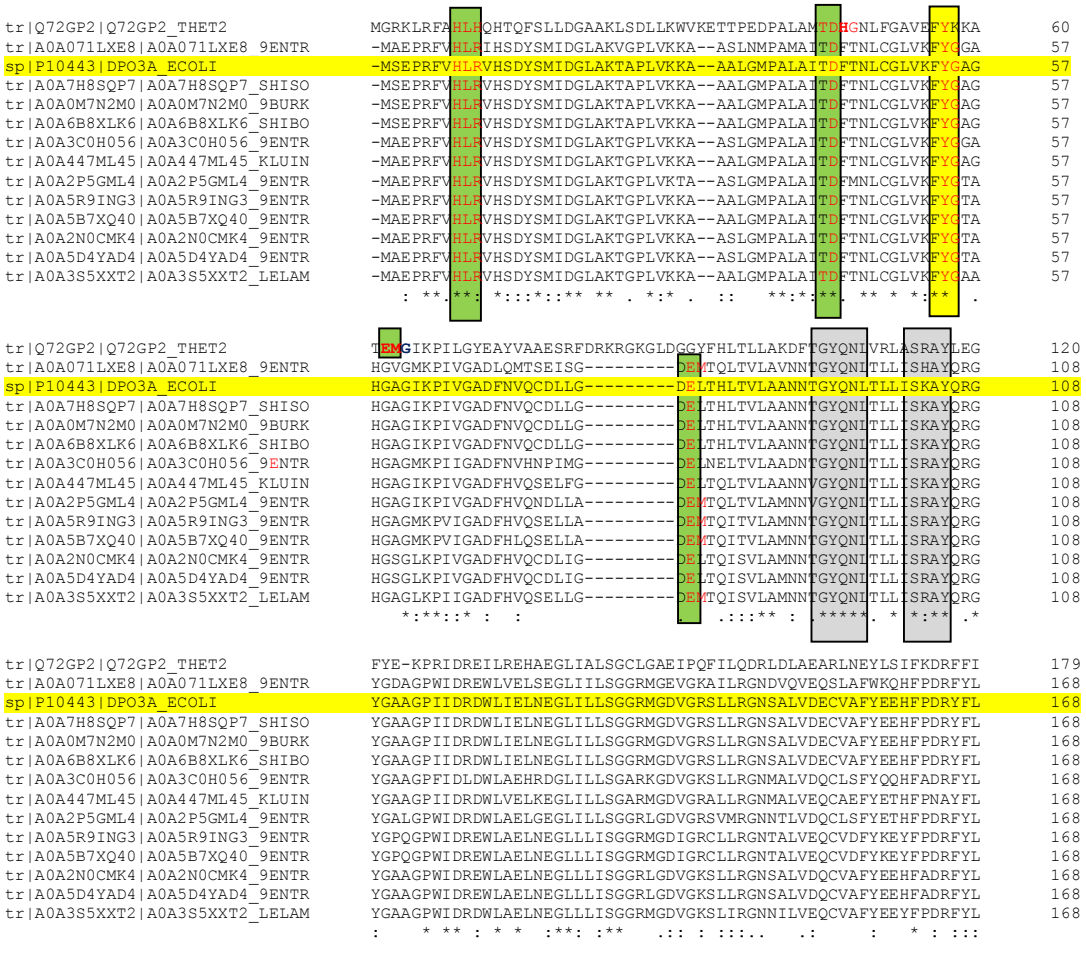
As shown in the earlier section, in the bacterial DNA polymerases III the PR function is performed by an independent subunit of the enzyme, (i.e.), by the ε-subunit which is a part of the MEC. In the MEC, the catalytic α-subunit of the polymerase III performs the polymerase function, whereas the ε-subunit performs the PR function. However, a novel PHP type PR exonuclease activity was reported in the DNA polymerase III α-subunit itself from the thermophilic bacterium, *Thermus thermophilus* [4]. They have reported that the PR exonuclease domain was in the same polypeptide as the polymerase domain and was located in the N-terminal region of the polymerase. They also found that the PR exonuclease was a typically a Zn^{2+} dependent enzyme, as the Zn^{2+} chelator *O*-phenanthroline inhibited the enzyme activity drastically even in the presence of 10 mM Mg^{2+}. Therefore, it was suggested that the functions of these two exonuclease activities could be complementary, i.e., the PHP enzyme might be more active on mismatches not preferred by the ε-exonuclease. From the MSA analysis it is clear that the active site of this type of enzyme consists of 4 motifs similar to other exonucleases of the PHP superfamily (Fig. 17). For example, the motif I has a dyad of histidines, which are separated by a single amino acid as HxH, which apparently coordinating the metal ion zinc. Motif II has an Asp/Glu, motif III has a His and motif IV has an Asp residue. It was proposed that the motifs II to IV might be involved in catalysis by participating in proton transfer and/or through metal ion coordination. The MSA shows the general active site pattern –HxH→E→H→D-, which is very similar to the pattern found in other PHP enzymes like DNA polymerases X and YcdX. The presence of the highly conserved template-binding YG pair (highlighted in yellow)

suggests that this PR function is template-dependent activity.

Mix and Match analysis of the α-subunit of DNA polymerases III of *E. coli* and other mesophilic enzymes along with the thermophilic counterpart is shown in Fig. 15 (only the N-terminal PHP domain is shown here). It shows that the mesophilic and thermophilic enzymes are not highly conserved but for the active site amino acids. However, some of the active site amino acids are not completely aligned in both the group of enzymes. Such a shift in the active site regions is already reported for invertases by Palanivelu [20]. Like other PHP enzymes this enzyme is also a strictly template–dependent enzyme as the template-binding YG pair is completely conserved in all (YK in the thermophilic enzyme) (highlighted in yellow). The metal-binding regions are highlighted in light green (Fig. 17).

However, the MSA shows that the conserved amino acids that are found in *T. thermophilus* enzyme are also found in other α-subunits of the mesophilic DNA polymerases III, but with slight modifications, e.g., **HLH→HL*R***; **D**HG→**D**FT; **EM**G→**EM**/LT; **DAR→DA*H*** (similar conserved sequences in mesophiles are shown in red) (Fig. 16). The first and the last triads are highly conserved with similar amino acids motif. These data suggest that the active site amino acids are still intact in the α-subunits of the DNA polymerases III in both the groups of organisms.

CLUSTAL O (1.2.4) Mix and Match analysis of the α-subunit of DNA polymerases III



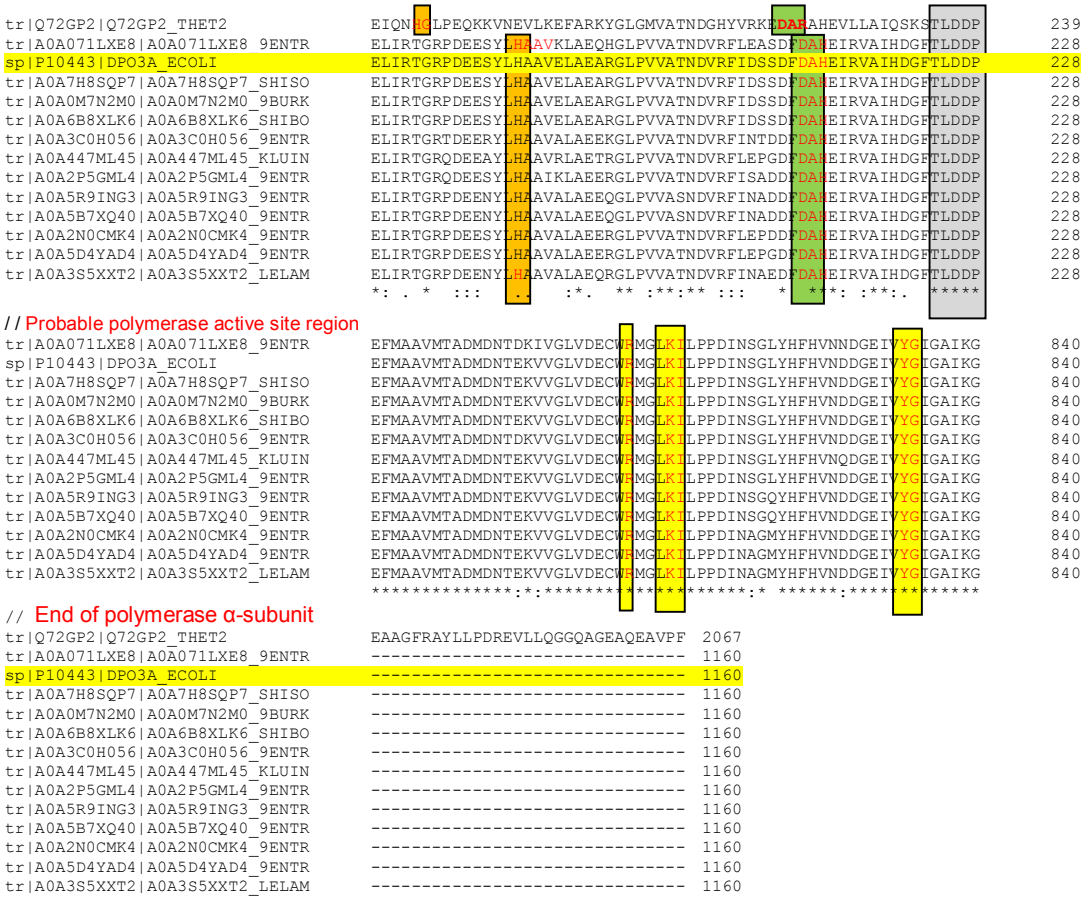


Fig. 17 Mix and Match analysis of the α-subunit of the thermophilic DNA polymerase III with mesophilic DNA polymerase III α-subunits

Q72GP2_THET2, Thermus thermophilus *A0A071LXE8_9ENTR, Mangrovibacter sp.*
P10443|DPO3A_ECOLI, Escherichia coli (K12) *A0A7H8SQP7_SHISO, Shigella sonnei*
A0A0M7N2M0_9BURK, Achromobacter sp. *A0A6B8XLK6_SHIBO, Shigella boydii*
A0A3C0H056_9ENTR, Enterobacteriaceae bacterium *A0A447ML45_KLUIN, Kluyvera intermedia*
A0A2P5GML4_9ENTR, Superficieibacter electus *A0A5R9ING3_9ENTR, Enterobacter sp.*
A0A5B7XQ40_9ENTR, Leclercia adecarboxylata *A0A2N0CMK4_9ENTR, Cedecea lapagei*
A0A5D4YAD4_9ENTR, Lelliottia nimipressurali; *A0A3S5XXT2_LELAM, Lelliottia amnigena*

Therefore, it is intriguing to know how *E. coli* and other mesophilic organisms have lost this activity or not detectable in their α-subunits even though both possess similar active site amino acids and motifs. Therefore, it is tempting to speculate that these mesophilic enzymes may also have the intrinsic co-editing function but possibly unexplored.

The T. *thermophilus* α-subunit of the DNA polymerase III and *B. subtilis* X polymerase shows – very similar active site patterns. For example, the ^{9}**H**L**H**-----^{46}**DH**G---^{61}**EM**G----^{220}**DA**R pattern at the very N-terminal region of the α-subunit of the DNA polymerase III is very similar to the ---^{339}**H**M**H**------^{370}**DH**S--^{428}**EM**D----^{526}**DA**H-pattern found in DNA polymerases X (numbering from *B. subtilis* DNA polymerase X) (Fig. 17).

5. PR FUNCTION IN CORONAVIRAL RNA POLYMERASES

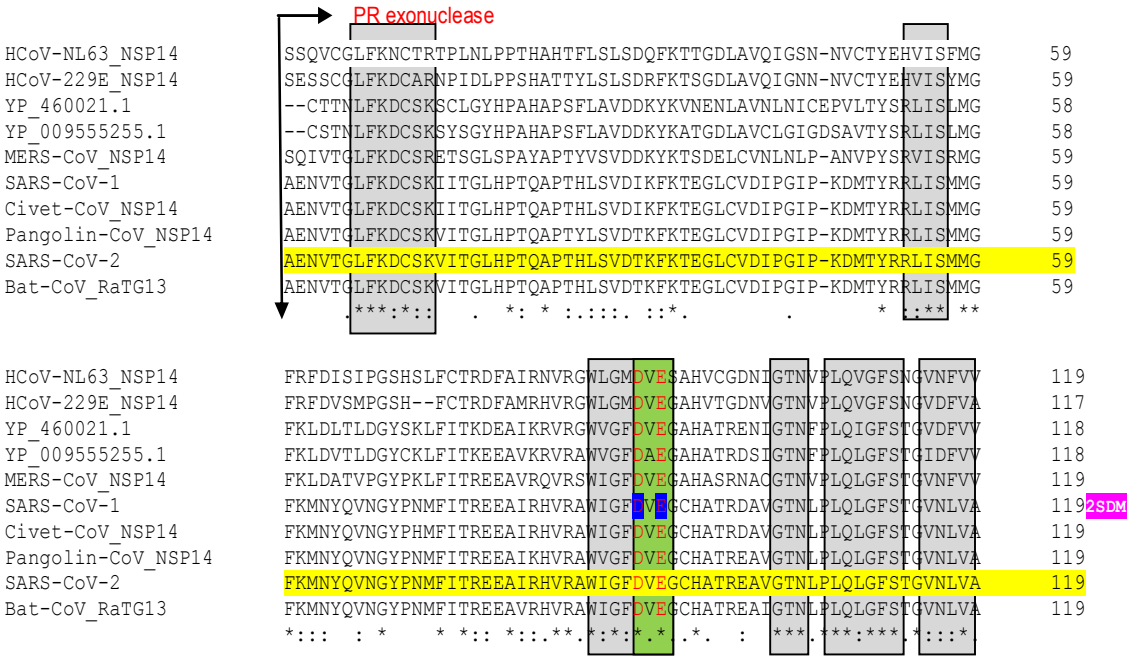
5.1 PR Function in RdRps of SARS-CoVs

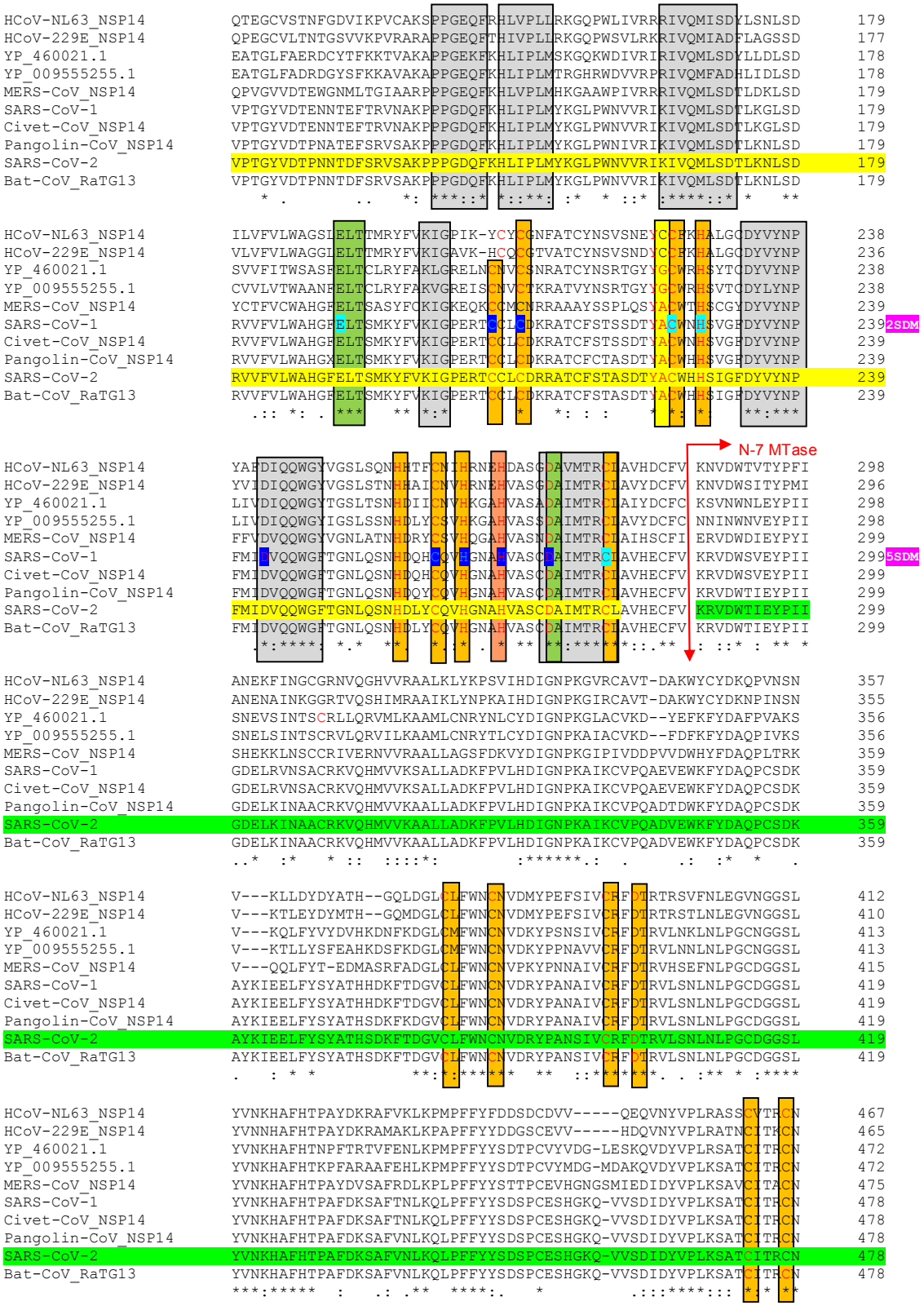
Majority of the human diseases are caused by RNA viruses [21]. All these RNA viruses replicate their genome as well as transcribe their genes using the single enzyme, the RdRp. Therefore, major advancements towards the design and development of antivirals are expected to come

from studies on the structure-function relationships of the enzymes of the RTC (RdRp, Primase, etc.) and PR exonuclease, also known as ExoN in CoVs. Active site analyses of the RdRps and Primases in SARS, SARS-related CoVs and HCoVs have been already reported by [22, 23 and references therein]. Therefore, in this communication, the ExoN which plays an important role in the replication of the viral genome is analysed to understand its active-site architecture and catalytic mechanism. As the ExoN of the SARS-CoVs directly interferes with the replication of the genome, it is also considered as a new target for drug design in addition to the RdRps and primases.

The ExoNs are found in nidoviruses only when their genome-size exceed beyond a certain threshold, i.e., >20 kb and thus, all nidoviruses with genome sizes above this threshold invariably possess the ExoN, whereas below the threshold do not possess the ExoN. [24]. Therefore, in all the four families of the *Nidovirales,* only the *Arteriviridae* (infects vertebrates) with a genome size ~15 kb do not possess the ExoN. The other three families, viz. *Coronaviridae* (~30 kb, infect vertebrates, mostly mammals), and the *Roniviridae* (~26 kb, infect mainly crustaceans) and *Mesoniviridae* (~20 kb, infect invertebrates) possess the Exon [25]. Therefore, the *Coronaviridae* with larger genome sizes, like the SARS-CoVs, SARS-related CoVs and HCoVs, invariably possess the ExoNs. It is suggested that the ExoNs might have been acquired by the three nidoviruses during evolution for the maintenance and stability of their larger genomes. The ExoN is encoded by the nonstructural protein 14 (NSP14) in CoVs and is found to be a bifunctional enzyme harbouring two enzyme activities on the same polypeptide, viz. a PR exonuclease activity at the N-terminal region (from amino acids 1 to 287), and a guanine-N7-methyltransferase (N7-MTase) activity for mRNA capping at the C-terminal region (from amino acids 288 to 527, numbering from SARS-CoV-1). Amino acids from 288 to 301 make a convoluted loop and a break in the loop resulted in the abolishment of the N7-MTase activity. The PR exonuclease and the N7-MTase domains perform PR function and viral mRNA capping, respectively, during genome replication and transcription processes. [26, 27] have found that these two domains function independently as in all ExoN knockout mutants the PR activity was severely affected, whereas the N7-MTase activity was not affected.

CLUSTAL O (1.2.4) MSA of ExoN and PR-exonucleases (Amino acids 1-287 represent PR exonuclease and 288-527 represent N7-MTase)




HCoV-NL63_NSP14
HCoV-229E_NSP14
YP_460021.1
YP_009555255.1
MERS-CoV_NSP14
SARS-CoV-1
Civet-CoV_NSP14
Pangolin-CoV_NSP14
SARS-CoV-2
Bat-CoV_RaTG13
N-7 MTase
2SDM
5SDM

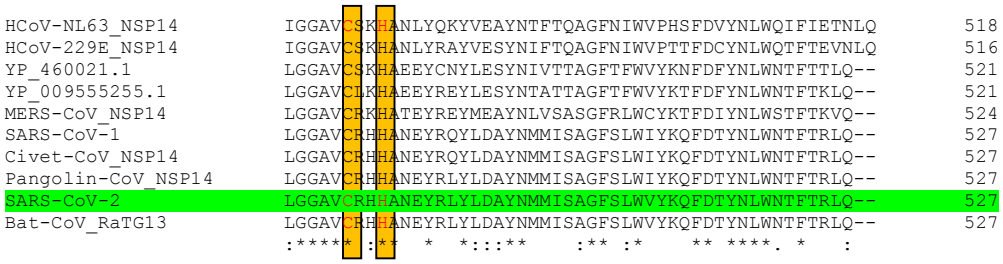


Fig. 18 MSA of the PR exonucleases from SARS, SARS-related CoVs and other HCoVs

Amino acids highlighted in light blue indicates the crystallographic data
JX104161.1 HCoV-NL63 (CBJ 037) *NC_002645, HCoV-229E*
YP_460021.1 HCoV-HKU1 *YP_009555255.1 HCoV-OC43*
NC_019843, MERS-CoV *NC_004718, SARS-CoV-1*
AAU04645, Civet-CoV *QIQ54047, Pangolin-CoV*
NC_045512.2, SARS-CoV-2 *MN996532.2, Bat (RaTG13)-CoV*

Furthermore, they also found that the SARS-CoV-2 ExoN knockout mutant was unable to replicate, suggesting a possibility for development of antivirals for ExoNs.

Fig. 18 shows the MSA of the ExoNs from SARS, SARS-related CoVs and HCoVs. The PR exonuclease region is highlighted in yellow and the MTase region is highlighted in green. But for few peptide regions, PR exonuclease regions are not highly conserved. Interestingly, the active site amino acids and the Zn binding motifs are completely conserved in all, suggesting their importance in the structure and function of these enzymes. The PR exonuclease domains of CoVs show that they all belong to the member of the DEDD superfamily of exonucleases [2]. The ExoN domains possess four possible Zn binding motifs (ZFMs) (two in the PR domain with the patterns-CxxC→CxxH- and -HxxxCxxH→C- and two in the N7-MTase domain with the patterns -Cx4C→CxxD- and –Cx3C→CxxH- (Fig. 18).

5.2 Active Site Analyses of the Exons of SARS-CoVs

The Exon's PR exonuclease active sites of SARS-CoV-1 and MERS-CoV and SARS-CoV-2 were studied by different investigators [28-29]. Minskaia et al [29] have analyzed the SARS-CoV-1 ExoN by SDM experiments and identified the residues D^{90}/E^{92} (motif I), D^{243} (motif II), and D^{273} (motif III) as the putative active-site residues. They have further demonstrated that modification of the ExoN active site amino acids resulted in the failure to recover infectious viral progenies. Ma et al [28] have analyzed the crystal structures of SARS-CoV-1 ExoN, in complex with its activator (NSP10) and functional ligands. They found that the amino acid residues Cys^{207}, Cys^{210}, Cys^{226}, and His^{229} constituted the first zinc finger whereas the second zinc finger was consisted of His^{257}, Cys^{261}, His^{264}, and Cys^{279}. Simultaneously mutating Asp^{90} and Glu^{92} to Ala, impaired the ExoN activity drastically, whereas E^{191}→A, H^{268}→A, or D^{273}→A mutants severely affected their ability to degrade RNA, confirming their importance in the exonuclease function. Asp^{243} is the fifth highly conserved amino acid identified in motif II. The ExoN activity of D^{243}→A mutant was completely lost, suggesting its possible role in the catalysis.

Ma et al. [28] have also found that the catalytic core of the SARS-CoV-1 ExoN was very similar to other DEDD superfamily of PR exonucleases but starkly differed from other PR exonucleases by the presence of two zinc fingers. Furthermore, SDM studies indicated that both these zinc fingers are essential for the function of the ExoN's PR function. For example, a set of mutations generated by SDM experiments on the PR domain of the ZFMs (C^{210}→H in ZF1 and C^{261}→A and H^{264}→R in ZF2) abolished replication of SARS-CoV-1 genome, suggesting the importance of both ZFMs in the genome viability of SARS-CoV-1. MSA shows that the 2 ZFMs (highlighted in orange) are highly conserved in all SARS, SARS-related CoVs and other HCoVs suggesting, one ZFM may play the structural role and the other one could possibly involve in catalysis (Fig. 19).

The MERS-CoV and SARS-CoV-2 ExoNs were analyzed by SDM experiments and knockout mutations by Ogando et al [27]. They subjected all the five predicted active-site amino acid residues of MERS-CoV's ExoN domain (D^{90}, E^{92}, E^{191}, D^{273}, and H^{268}) by replacing them with Ala or with more conservative substitutions like D to

E or Q; E to D or Q. This SDM experiment yielded a total of 14 ExoN active-site mutants, including the D^{90}→A/E^{92}→A (motif I), double mutant, which was frequently used as a prototypic viable ExoN knockout mutant in SARS-CoV-1 studies. The following SDM experiments of the MERS-CoV ExoN yielded non-viable phenotypes: D^{90}→A/Q/E; E^{92}→A/D/Q; E^{191}→A/Q; D^{273}→A/E/Q; H^{268}→A, suggesting their importance in the activity of the PR exonuclease. Furthermore, the ZF mutations (C^{210}→H in ZF1 and C^{261}→A and H^{264}→R in ZF2) abolished MERS-CoV replication, further establishing the importance of both the ZFMs for MERS-CoV viability. These results are also in close agreement with the SARS-CoV-1 results obtained by Ma et al [28]. They further evaluated the impact of ExoN inactivation (using a D^{90}→A/E^{92}→A ExoN motif I double mutant) on SARS-CoV-2 replication and viability. Surprisingly, they could not rescue any viable progeny in which the two key residues of the ExoN active site amino acids were mutated. Interestingly, all ExoN exonuclease knockout mutations that proved lethal in reverse genetics were found to severely decrease ExoN activity without affecting N7-MTase activity. The SDM and crystallographic analyses of the PR exonuclease domains of the ExoNs of SARS-CoV-1, MERS-CoV and SARS-CoV-2 have clearly established that the same set of amino acids are making the active site as shown in Fig. 19.

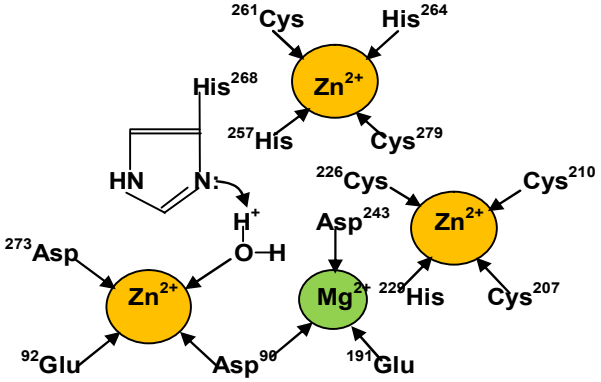


Fig. 19 Proposed amino acids at the active site for the PR exonuclease of SARS-CoV-2

Table 1 Summary of the PR 3'-5' exonucleases from different bacteria and CoVs

Family	Consensus As* Pattern	Proton Acceptor	Catalytic Metal ion**	Zn-Binding site(s)
DEDD				
DNA Pol I	-DxE-D-Y-D-	Tyr	Zn^{2+}	1
DNA Pol II	-DxE-D-Y-D-	Tyr	Zn^{2+}	1
RNase D	-DxE-D-Y-D-	Tyr	Zn^{2+}	1
DNA Pol III, ε-subunit	-DxE-D-H-D-	His	Zn^{2+}	1
RNase T	-DxE-D-H-D-	His	Zn^{2+}	1
^SARS-CoV-1 ExoN/ACE2$	-DxE-D-H-D-	His	Zn^{2+}	3
MERS-CoV ExoN/DPP4#	-DxE-D-H-D-	His	Zn^{2+}	3
SARS-CoV-2 ExoN//ACE2	-DxE-D-H-D-	His	Zn^{2+}	3
HCoV-NL63 ExoN/ACE2	-DxE-D-H-D-	His	Zn^{2+}	3
PHP				
DNA Pol X (*B. Subtilis*)	-HxH-E-H-D-	His	Zn^{2+}	1
YcdX (*E. coli*)	-HxH-E-H-D-	His	Zn^{2+}	3
DNA Pol III (*Tth*) co-editing	-HxH-E-H-D-	His	Zn^{2+}	1

**As, Active site; Pol, polymerase*
***proposed catalytic metal ions, water bound Zn^{2+}*
^Similar active sites and structural features are found in SARS-related CoVs and HCoVs too
$ACE2, Angiotensin-Converting Enzyme 2; #DPP4, Dipeptidyl peptidase 4

Table 1 summarizes the minimum participating amino acids in both the superfamilies of exonucleases. The first two amino acids are acidic (DxE) in DEDD superfamily, whereas they are basic (HxH) in PHP superfamily. The next three amino acids are functionally equivalent in both.

5.3 Role of ExoN in the Functioning of the RdRp in CoVs

The ExoN activity is stimulated by binding of the NSP10. The activated ExoN is now placed, likely next to the RdRp's polymerization site during replication. As soon as the elongation starts, the N7-MTase caps the RNA to mimic the host mRNAs and to avoid degradation by host exonucleases. By a mechanism of stuttering at the slippery stop-site present at the end of viral genome, it adds the poly-A chain and thus, making the genome replication process complete. Only this faithfully replicated RNA, also known as the genomic RNA, is encapsidated and assembled into virions (Fig. 20).

Even though the replication and transcription processes are accomplished by the same set of enzymes, mechanistically both appears to be different. For example, the viral the genomic RNA is replicated by a continuous process from end to end, but the transcription process is not a continuous process, but a discontinuous process and a complex one. For example, the SARS-CoV-2 makes 1 genomic (1ab) and 9 subgenomic mRNAs (S, 3a, E, M, 6, 7a, 8, 7b and N, 10?). All the mRNAs are capped at their 5'-ends and tailed at their 3'-ends like the host mRNAs for subsequent translations. Invariably, all the genomic and the subgenomic RNAs in CoVs contain a transcription-regulating sequence (TRS) at their 5'-ends, which are located immediately adjacent to the ORFs. Each subgenomic mRNA contains the common 5-''leader'' sequence (~70 nt) fused to the ''body'' TRS also known as TRS-B. The TRS-L and TRS-B sites have a conserved core sequence (CS) of 7 to 8 nt (-AACGAAC- is the CS in SARS-CoV-2). Thus, the genomic and subgenomic RNAs have the same common leader sequence and TRS at their 5'-ends. Because TRS-B is a signal for RdRp to switch templates, it is possible that recombination events are more likely to occur at or near TRS-B sites. As mentioned elsewhere, all the subgenomic mRNAs are created by discontinuous transcription. The discontinuous transcription requires base-pairing between *cis*-acting body TRSs, with the leader sequence located at the 5′-end of the viral genome. Because of such discontinuous extension of minus strands, all these subgenome-length minus strands carry the complement of the leader sequence at their 3'-ends [29, 30]. *During replication mode, the Replication-Transcription Complex (RTC) ignores the transcription signals.*

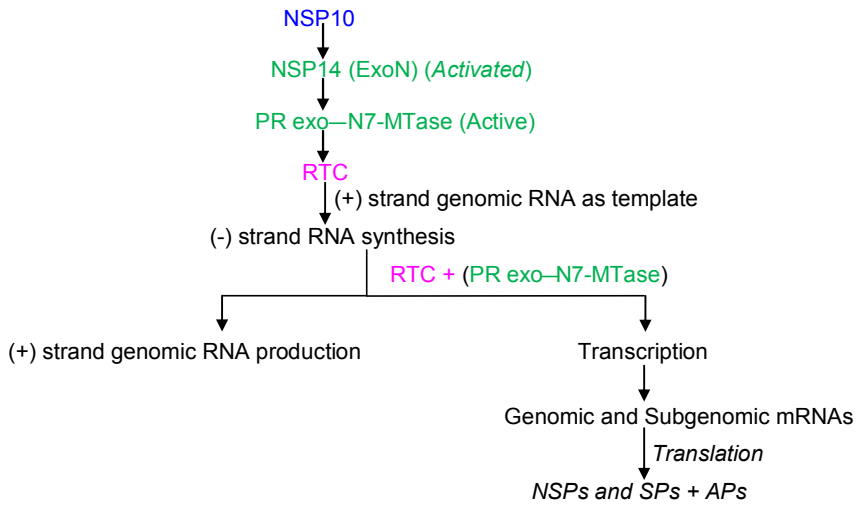


Fig. 20 A proposed simplified model for the replication, transcription and translation processes in SARS-CoVs

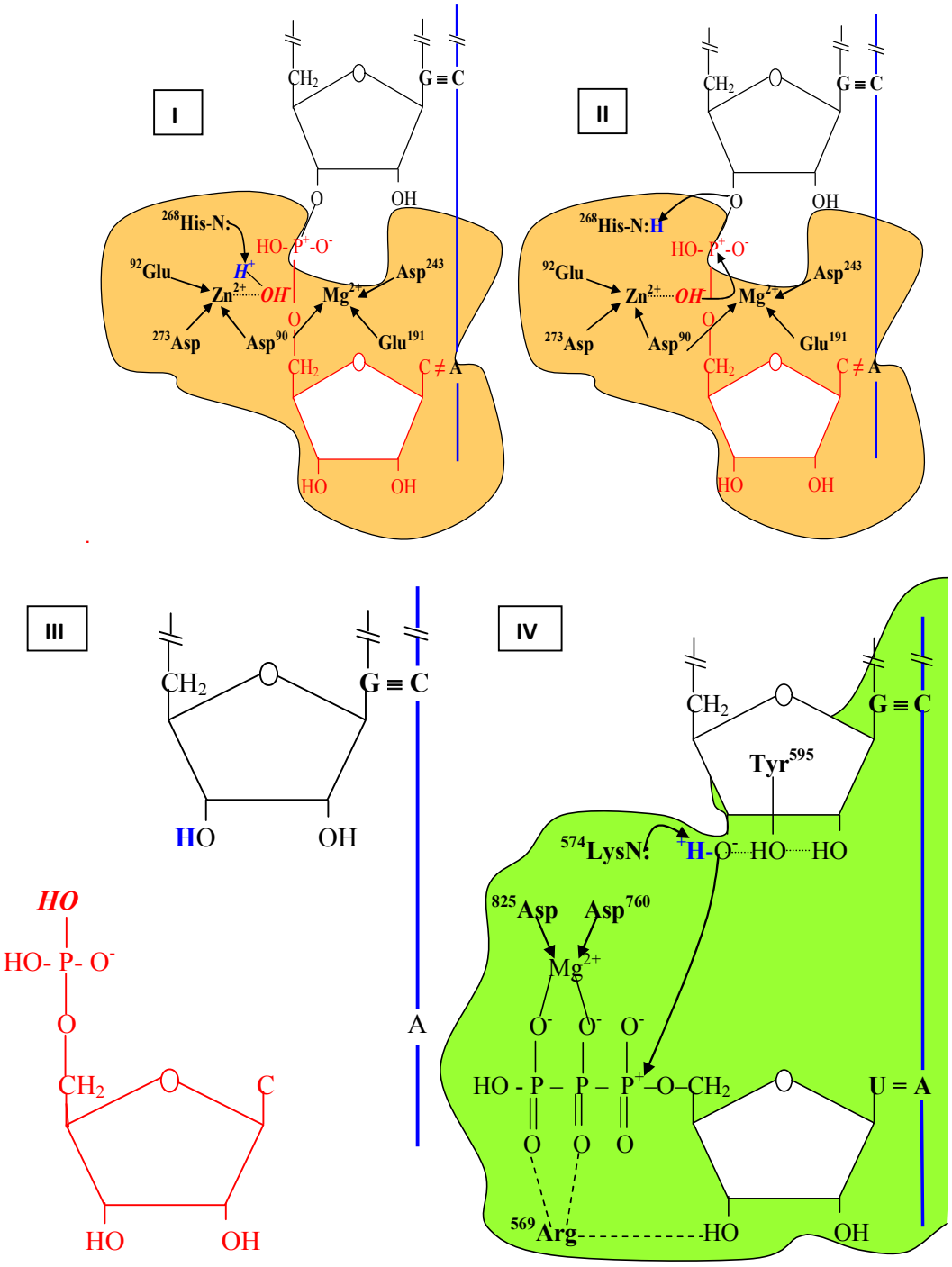


Fig. 21 Steps (1-4) involved in the proposed mechanism of action of ExoN PR Exonuclease of SARS-CoV-2

Steps 1 and 2: *The wrongly added base is excised by the PR exonuclease by the mismatch induced activation of the water molecule bound to Zn^{2+}, initiating proton transfer with the simultaneous nucleophilic attack on the susceptible phosphodiester bond by the highly reactive Zn-hydroxide.*

Step 3 and 4: *The wrongly added base is excised, and the polymerase resume synthesis with the right nucleotide*

The RTC plays three different roles during replication and transcription processes, viz. i) synthesis of the (-) RNA strand using genomic (+) strand as template, ii) production of large number of the (+) genomic RNAs from the (-) RNA strand as template and iii) transcription of genomic and subgenomic mRNAs from the (-) RNA strand. The mutations that occur during the (-) strand synthesis will be reflected in the (+) strand genomic RNAs and in the genomic and subgenomic mRNAs and hence, in the nonstructural, structural and accessory proteins (NSPs, SPs, APs). Effecting a large deletion, the 5" leader sequence is fused to each subgenomic RNAs, which ensures discontinuous transcription of subgenomic mRNAs. Fig. 20 shows a proposed model for the replication, transcription and translation processes in SARS-CoVs.

6. GENERAL MECHANISM PROPOSED FOR PR EXONUCLEASES IN DNA/RNA SYNTHESIS AND MODIFICATIONS

A two-metal ion mechanism is proposed for the PR exonucleases from both the superfamilies. A water-bound Zn^{2+} is proposed as the primary metal ion which initiates the catalysis and the Mg^{2+} which is also making the active site, is proposed as the supporting metal ion. The mismatch at the active site of the polymerase, signals the PR exonuclease to move-in and excise the mismatch. Zinc is placed as the primary metal ion as it is one of the most ubiquitous cofactors found in a large number of enzymes and proteins. For example, >450 enzymes and proteins use Zn^{2+} as cofactor. In these enzymes, the zinc atoms are known to play both the structural and catalytic roles. For example, zinc based catalysis is established in many enzymes like carboxypeptidases-A, carbonic anhydrases, thermolysin, alkaline phosphatases, metallo β-lactamases, PR exonucleases of DNA and RNA polymerases, RNA modifying enzymes, etc. [31 and references therein]. One common theme proposed for many of these enzymes is the activation of a water molecule coordinated to the Zn^{2+} for a nucleophilic attack on the carbonyl carbon of a peptide bond, or the phosphorus atom of a phosphoester bond. Furthermore, it is proposed that several of the zinc enzymes' action is facilitated by the formation of a zinc-hydroxide [1 and references therein]. During instances of mismatch of nucleotides, the RdRp stalls/pause, allowing the PR domain to excise the mismatched nucleotide. Fig. 21 shows a proposed mechanism for PR exonuclease function for excising a mismatched nucleotide during RNA replication in SARS-CoV-2.

In addition to the catalytic zinc, a second active site metal ion like Mg^{2+}/Zn^{2+} is also known to be essential for assisting the catalysis by SDM experiments. The secondary metal ion is suggested to function to stabilize the transient pentacovalent species and/or to facilitate the leaving of the 3' oxyanion from an axial position, whereas the primary metal ion facilitate the formation of an attacking hydroxide ion. Involvement of both the metal ions has been unequivocally proved by X-ray crystallographic data and SDM experiments. The PHP superfamily of PR exonucleases invariably use Zn^{2+} as the primary metal ion, as other metal ions could not fit into its place whereas the DEDD superfamily of enzymes which use Zn^{2+} as the primary metal ion, but other divalent metal ions like Mg^{2+}/Mn^{2+} could also fit into the primary Zn^{2+} site as it is essentially made up of three acidic amino acids [32]. In the DEDYD subfamily of exonucleases, Tyr is used as the proton acceptor. In fact, Tyr serves as the nucleophile in the active site of topoisomerases, which also makes phosphodiester bond breaks during catalysis [33].

7. CONCLUSIONS

The PR functions form an important component of the replicative polymerases in biological systems. This study shows there are at least two different types of PR exonucleases performing this function in prokaryotes and viruses (CoVs). They belong to either DEDD or PHP superfamilies. The DEDD superfamily of exonucleases essentially uses the four acidic amino acids, DEDD in their catalytic site with additional Y or H as the proton acceptor (-DEDDY/H-). The PHP superfamily essentially uses -HxH-E-H-D- pattern at the active site with an invariant H as the proton acceptor. The PHP exonucleases are found to be mainly a Zn^{2+}-dependent enzymes. From the similarities of active site amino acids/motifs, it may be concluded that the DEDD and PHP superfamilies of PR exonucleases should have evolved from a common ancestor but diverged very long ago. The structural features of the PR enzymes from the CoVs suggest that CoVs may have acquired the exonuclease function, possibly from a prokaryote. However, the presence of two zinc-binding sites in PR active site of SARS, SARS-related CoVs and HCoVs sets it apart from their homologues.

ACKNOWLEDGMENTS

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

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. 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, Vol 1, pp 1-39, ISBN: 9788193422441. DOI: 10.9734/bpi/rabr/v1; 2019.
2. Zuo Y, Deutscher MP. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. Nucleic Acids Res. 2001;29:1017–1026.
3. Nagpal S, Nair DT. The PHP domain of PolX from *Staphylococcus aureus* aids high fidelity DNA synthesis through the removal of misincorporated deoxyribo-, ribo- and oxidized nucleotides. Nat Sci Rep. 2021;11:4178.
4. Stano NM, Chen J, McHenry CS. A co-proofreading Zn^{2+}-dependent exonuclease within a bacterial replicase. Nat Struct Mol Biol. 2006;13:458–459.
5. Teplyakov A, Obmolova G, Khil PP, Howard AJ, Camerini-Otero RD, Gilliland GL. Crystal Structure of the *Escherichia coli* YcdX Protein Reveals a Trinuclear Zinc Active Site. Proteins: Struct Funct Genet. 2003;51:315–318.
6. Derbyshire V, Freemont PS, Sanderson MR, Beese L, Friedman JM, Joyce CM, Steitz TA. Genetic and crystallographic studies of the 3,5'-exonucleolyitc site of DNA polymerase I. Science. 1988;240: 199-201.
7. Derbyshire V, Grindley ND, Joyce CM. The 3'-5' exonuclease of DNA polymerase I of Escherichia coli: contribution of each amino acid at the active site to the reaction. EMBO J. 1991;10:17-24.
8. Joyce CM, Steitz TA. DNA polymerase I: from crystal structure to function via genetics. Trends Biochem Sci. 1987;12: 288-292.
9. Beese LS, Steitz TA. Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. EMBO J. 1991;10:25–33.
10. Wang F, Yang W. Structural insight into translesion synthesis by DNA Pol II. Cell. 2009;139:1279–1289.
11. Fijalkowska IJ, Schaaper RM. Mutants in the Exo I motif of *Escherichia coli* dnaQ: Defective proofreading and inviability due to error catastrophe. Proc Natl Acad Sci. (USA). 1996;93:2856-2861.
12. Cisneros GA, Perera L, Schaaper RM, Pedersen LC, London RE, Pedersen LG, Darden TA. Reaction mechanism of the *ε* subunit of *E. coli* DNA polymerase III: Insights into active site metal coordination and catalytically significant residues. J Am Chem Soc. 2009;131:1550–1556.
13. Hamdan S, Carr PD, Brown SE, Ollis DL, Dixon NE. Structural Basis for Proofreading during Replication of the *Escherichia coli* Chromosome. Structure. 2002;10:535–546.
14. Blanco L, Bernad, A, Salas M. (1992). Evidence favouring the hypothesis of a conserved 3'-5' exonuclease active site in DNA-dependent DNA polymerases. Gene. 1992;112:139-44.
15. Zuo Y, Zheng H, Wang Y, Chruszcz M, Cymborowski M, Skarina T, Savchenko A, Malhotra A,Minor W. Crystal Structure of RNase T, an exoribonuclease involved in tRNA maturation and end-turnover. Structure. 2007;15:417–428.
16. Ramadan K, Shevelev I, Hübscher U. The DNA-polymerase-X family: controllers of DNA quality?. Nat. Rev. Mol. Cell Biol. 2004;5:1038–1043.
17. Uchiyama Y, Takeuchi R, Kodera H, Sakaguchi K. Distribution and roles of X-family DNA polymerases in eukaryotes. Biochimie. 2009;91:165–170.
18. Banos B, Lazaro JM, Villar L, Salas M, de Vega M.Editing of misaligned 3'-termini by an intrinsic 3'-5' exonuclease activity residing in the PHP domain of a family X DNA polymerase. Nucleic Acids Res. 2008;36:5736-5749.
19. Nakane S, Nakagawa N, Kuramitsu S, Masui R. Characterization of DNA polymerase X from *Thermus thermophilus* HB8 reveals the POLXc and PHP domains are both required for 3'–5' exonuclease activity. Nucleic Acids Res. 2009;37:2037–2052.

20. Palanivelu P. Multiple sequence analysis of Polygalacturonases and Invertases and Phase Shift in Conserved Motifs. Ind J Biotechnol. 2007;6:24-30.
21. Poltronieria P, Sunb B, Mallardoc M. RNA Viruses: RNA Roles in Pathogenesis, Coreplication and Viral Load. Current Genomics, 2015;16:327-335.
22. Palanivelu P. Analyses of the Spike Proteins of Severe Acute Respiratory Syndrome-Related Coronaviruses. Microbiol Res J Int. 2020;30:32-50.
23. Palanivelu P.RNA-Dependent RNA Polymerases of Severe Acute Respiratory Syndrome-Related Coronaviruses- An Insight into their Active Sites and Mechanism of Action. International J Biochem Res Rev. 2021;29:29-52.
24. Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ (2006) Nidovirales: Evolving the largest RNA virus genome. Virus Res. 2006;117:17–37.
25. Lauber C, Goeman JJ, Parquet MDC, Nga PT, Snijder EJ, Morita K, Gorbalenya AE. 2013. The footprint of genome architecture in the largest genome expansion in RNA viruses. PLoS Pathog Available:9:e1003500.https://doi.org/10.1371/journal.ppat.1003500.
26. Ogando NS, Ferron F, Decroly E, Canard B, Posthuma CC, Snijder EJ. The Curious Case of the Nidovirus Exoribonuclease: Its Role in RNA Synthesis and Replication Fidelity. Front Microbiol. 10:1813. DOI: 10.3389/fmicb.2019.01813.
27. Ogando NS, Zevenhoven-Dobbe JC, Meer YVD, Bredenbeek PJ, Posthuma CC, Snijder EJ. The Enzymatic Activity of the nsp14 Exoribonuclease Is Criticalfor Replication of MERS-CoV and SARS-CoV-2. J Virol. 2020;94:e01246-20.
28. Ma Y, Wu L, Shaw N, Gao Y, Wang J, Sun Y, Lou Z, Yan L, Zhang R, Rao Z. Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. Proc Natl Acad Sci (USA). 2015;112:9436 –9441.
29. Minskaia E, Hertzig T, Gorbalenya AE, Campanacci V, Cambillau C, Canard B, Ziebuhr J. Discovery of an RNA virus 3'-5' exoribonuclease that is critically involved in coronavirus RNA synthesis. Proc Natl Acad Sci (USA). 2006;103:5108–5113.
30. Yang Y, Yan W, Hall AB, Jiang X. Characterizing Transcriptional Regulatory Sequences in Coronaviruses and Their Role in Recombination. Mol Biol Evol. 2021;38:1241–1248.
31. Coleman JE. Zinc enzymes. Curr Opin Chem Biol. 1998;2:222–234.
32. Freemont PS, Friedman JM, Beese LS, Sanderson MR, Steitz TA. Proc Natl Acad Sci. (USA). 1988;85:8924-8928.
33. Champoux JJ. DNA Topoisomerases: Structure, Function, and Mechanism. Annu Rev Biochem. 2001;70:369–413.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle4.com/review-history/70674