Univerzita Karlova

Přírodovědecká fakulta

Studijní program: Speciální chemicko-biologické obory

Studijní obor: Molekulární biologie a biochemie organismů

Kristýna Pokorná

Netemplátové aktivity DNA/RNA polymeráz a jejich význam

Non-template activities of DNA/RNA polymerases and their significance

Bakalářská práce

Školitel: Mgr. Václav Vopálenský, Ph.D.

Praha, 2023

Poděkování

Ráda bych poděkovala svému školiteli, Mgr.Václavu Vopálenskému, Ph.D, za ochotu, trpělivost, vstřícnost a smysl pro humor při spolupráci, dále své rodině za morální a finanční podporu.

Tato práce vznikla za podpory projektu Národní institut virologie a bakteriologie (Program EXCELES, ID: LX22NPO5103), financovaného Evropskou unií – Next Generation EU.

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 3. května 2023

Kristýna Pokorná

Abstrakt:

Netemplátové přidávání adenosinu na 3´ konec produktu u některých DNA polymeráz často využívaných v molekulární biologii je dosud neobjasněný fenomén, se kterým se potýkáme již několik desetiletí. Existuje jen málo jiných příležitostí, při kterých dochází k netemplátovému přidávání adenosinu na 3´ konec informační molekuly v buňce, a jednou z nich je právě polyadenylace. Zároveň je známo, že určité RNA viry vykazují podobné aktivity v rámci polyadenylace či ko-transkripční modifikace svých mRNA. Tato práce se zaměřuje na detaily těchto tří procesů, na vztahy mezi nimi a dochází k závěru, že jsou všechny odvozeny ze stejné konzervované vlastnosti ancestrální polymerázy.

Klíčová slova: DNA polymeráza, RNA polymeráza, netemplátová syntéza, Taq polymeráza, polyadenylace

Abstract:

Non-templated addition of adenosine on the 3´ end of the product of several DNA polymerases commonly used in molecular biology has been a problematic yet unexplained phenomenon for a few decades now. The only other instance where adenosine is added to the 3´ end of an information molecule in a cell is polyadenylation. At the same time, several RNA viruses appear to perform similar action to either edit or polyadenylate their mRNAs. This work focuses on the details of these three highlighted processes and relationship between them and has found out that all of them appear to stem from a common conserved property of the ancestral polymerase.

Key words: DNA polymerase, RNA polymerase, non-templated addition, Taq polymerase, polyadenylation

Obsah:

Seznam použitých zkratek :

zkratka	CZE	ENG
Å	Angstrom	Angstrom
A	Adenin	Adenine
Approx.	zhruba	approximately
AMP	adenosin monofosfát	adenosine monophosphate
ATP	adenosin trifosfát	adenosine triphosphate
C	Cytosin	cytosine
CR	konzervovaný region	conserved region
C-terminal	C-konec	C- end
dATP	deoxyadenosin trifosfát	deoxyadenosine triphosphate
dITP	deoxyinositol trifosfát	deoxyinositol triphosphate
DNA	deoxyribonukleová kyselina	deoxyribonucleic acid
E. coli	*Escherichia coli*	*Escherichia coli*
G	Guanine	Guanine
kDa	kilodalton	kilodalton
LUCA	poslední univerzální společný předek	last universal common ancestor
Mg^{2+}	hořečnatý kationt	magnesium cation
mRNA	mediátorová RNA	messenger RNA
N - terminal	N-konec	N-end
PAP/PAPase	poly (A) polymeráza	poly (A) polymerase
PCR	polymerázová řetězová reakce	polymerase chain reaction
PNP/PNPase	polynukleotid fosphoryláza	polynucleotide phosphorylase
PPi	pyrofosfát	pyrophosphate
RdRp	RNA dependentní RNA polymeráza	RNA dependent RNA polymerase
RNA	ribonukleová kyselina	ribonucleic acid
ssRNA	jednovláknová RNA	single stranded RNA
T	Thymin	Thymine
tRNA	transferová RNA	transfer RNA
U	Uridin	Uridine
UTR	nepřekládaná oblast	untranslated region

Introduction

Thermus aquaticus DNA polymerase I (Taq polymerase) was first isolated by Chien et al. in 1976 from a thermophilic eubacterium *Thermus aquaticus* and eventually replaced Klenow fragment of DNA polymerase I from *Escherichia coli* in Polymerase Chain Reaction (PCR) that was used previously, due to its tolerance for higher temperatures (Saiki et al., 1988; Chien et al., 1976). However, it became apparent that Taq polymerase, as well as several others, sometimes add non-templated extra nucleotide to the 3′ stand end (Clark, 1988; Hu, 1993; Fiala et al, 2007). This phenomenon is mostly undesirable as it creates errors in genotyping but has proven useful for T/A cloning. Currently, the reason as to why do the polymerases behave this way is unknown. At the same time, there are factors that influence the probability of a non-templated extra nucleotide addition in Taq polymerase, such as the sequence of the 5′ reverse primer or adjacent nucleotide (Brownstein et al., 1996), therefore chances are that the observed phenomenon is not a matter of random error activity. Also, the non-templated nucleotide added by Taq polymerase is predominantly adenosine preferably added next to cytosine (Brownstein et al., 1996; Hu, 1993).

Being predominantly adenosine added to the 3′ end strand (Brownstein et al., 1996; Hu, 1993), one is inclined to find similarities between the process of nuclear polyadenylation of pre-mRNA molecules and the non-templated nucleotide addition in Taq polymerase. Polyadenylation, the addition of adenosines to the 3′ end of a pre-mRNA during nuclear post-transcription modifications (Rorbach et al., 2014) and in other circumstances are usually performed by two different enzymes – poly (A) polymerase and polynucleotide phosphorylase (Yehudai-Resheff et al., 2001; Sarkar and Fisher, 2006; Edmonds and Abrams, 1960). Nevertheless, that is not always the case in viruses – certain, mostly negative strand RNA viruses contain an enzyme RNA dependent RNA polymerase that performs multiple functions such as RNA synthesis, editing and polyadenylation of the 3′ end of product – for example *Paramyxovirus* (Jordan et al., 2018).

Therefore the purpose of this work will be to assess whether, based on the current available research, it is possible to explain the non-templated adenosine addition in Taq polymerase as being a rudiment of an ancestral polyadenylating function of the enzyme with the aid of RNA dependent RNA polymerase function in several RNA viruses.

1. Polyadenylation in bacteria and Taq polymerase

Polyadenylation, the process of non-templated synthesis of mostly adenosine monophosphate sequences at the 3´ end of mRNA, can be encountered in almost every single living organism (Rorbach et al., 2014; Zhuang et al., 2013; Slomovic et al., 2006; Mohanty and Kushner, 2010). The only exceptions are organisms such as *Mycoplasma gallisepticum* and several Archaea, eg. *Haloferax* (Portnoy et al., 2005) who apparently lost the ability secondarily in evolution (Portnoy and Schuster, 2008).

The resulting structure from polyadenylation is referred to as a poly (A) tail, despite also containing other nucleotides on occasion. Therefore poly (A) tails consisting strictly out of adenosine monophosphates are labelled as homopolymeric while those that have another nucleotide incorporated (yet majority still being adenosine monophosphate) are heteropolymeric (Rorbach et al., 2014; Slomovic et al., 2006).

Apart from nuclear polyadenylation occurring in nucleus as a part of post-transcription modifications, there is also another type that occurs in the cytoplasm of eukaryotes – cytoplasmic polyadenylation. Discovered in the 1970´s (Villalba et al., 2011) in embryonic development and oocytes (Charlesworth et al., 2013), it is a process of re-activation of de-activated mRNAs by modifying their poly (A) tails in the cytoplasm (Villalba et al., 2011; Cui et al., 2013) by shortening their poly (A) tail to under 20 nucleotides from the standard 100-200 nucleotide long tails and then, once reactivated, being lengthened to 80-150 nucleotides (Richter, 1999). Aside from already mentioned key role in spermatogenesis investigated in *Caenorhabditis elegans* (Luitjens et al., 2000) and oogenesis in *Drosophila* (Cui et al., 2013), it has been also found to up-regulate the development of new neural synapses in *Aplysia* (Liu and Schwartz, 2003).

Nuclear polyadenylation in eukaryotes has been thoroughly studied (Zhuang et al., 2013; Colgan and Manley, 1997), yet in the case of prokaryotic cell type, much less is known about the process (Rorbach et al., 2014; Mohanty and Kushner, 2010; Slomovic et al., 2006). Generally, it is apparent that polyadenylation in prokaryotic cell types serves a very different purpose than in eukaryotes – its primary role is being a degradation signal rather than a degradation prevention and stability increasing component of the mRNAs like in eukaryotes (O´Hara et al., 1995; Rott et al., 2003). The mechanism begins by endonucleolytic cleaving, synthesis of poly (A) tail and degradation of the mRNA. This also applies to endosymbiotic organelles of eukaryotic cells (Slomovic et al., 2006). However in organelles, polyadenylation

and degradation is done by a single enzyme called polynucleotide phosphorylase. Prokaryotic cell types also have a poly (A) polymerase which organelles lack (Yehudai-Resheff et al., 2001). Next, bacterial polyadenylate tails are generally shorter, about 30 nucleotides (Anantharaman et al., 2002; O´Hara et al., 1995), than eukaryotic that are usually longer than 100-200 nucleotides (Richter, 1999).

1.1 Polynucleotide phosphorylase

Polynucleotide phosphorylase (PNP) was discovered in the 1950´s as an RNA polymerase (Grunberg-Manago et al., 1955). This enzyme has the ability to perform two enzymatic activities – either act as a 3´ untemplated ssRNA polymerase or as a 3´ to 5´ exonuclease (Rorbach et al., 2014; Mohanty and Kushner, 2000). The polymerase activity synthesizes short heteropolymeric poly (A) tails (approx. 83% A, 5% U, 5% C, 7% G), which is more profound once poly (A) polymerase is absent in the *E.coli* model as well as its abundance in the cell (Mohanty and Kushner, 2000). The exonuclease activity is apparently influenced by the strength of the secondary structure of the mRNA – if the secondary structure of the template is weak, polynucleotide phosphorylase is enticed to cleave it while if it encounters a strong secondary structure, it does not proceed to cleavage (Rorbach et al., 2014). If the concentration of inorganic phosphate is high in the cytoplasm, the PNPase degrades the poly (A) tail while if the concentration of inorganic phosphate is low, it synthesizes them (Figure 1).

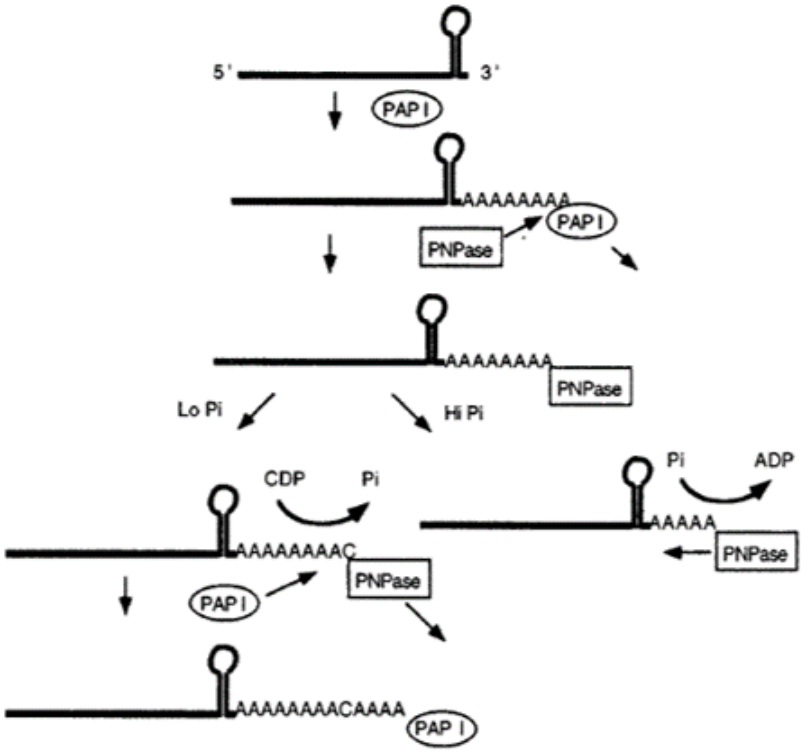

Figure 1 - *PNPase mechanism of function - in E.coli, poly (A) polymerase (PAP1) mainly performs polyadenylation of the 3′ end of the pre-mRNA while polynucleotide phosphorylase (PNPase) either cleaves the poly (A) tail if the cytoplasmic concentration of phosphates (Pi) is high or synthetises short heteropolymeric tails if the concentration of phosphates is low (Mohanty and Kushner, 2000)*

Regarding the structure, PNPase is a trimer 85 Å large with a central channel tunnel in the centre, molecular weight of 30 000 (Valentine et al., 1969), consisting of two core domains, a helical domain on the lower core domain (Symmons et al., 2000), S1 domain (Bycroft et al., 1997) and KH domain in the upper part of the core domain (Gibson et al., 1993; Garcia-Mena et al., 1999). The first core domain is active in RNA degradation, the second core domain degrades on polyadenylated mRNA and the S1 domain contains the binding site for the poly (A) tails (Yehudai-Resheff et al., 2003; Régnier et al., 1987). The PNPase trimer is held together by hydrophobic interactions of the B-sheets (Figure 2) (Symmons et al., 2000). The enzyme works by physically sticking to a substrate that is located inside the channel about 9 Å wide which is the same principle as in DNA/RNA polymerases however PNPase is not considered to be homologous to them as it differs topologically (Nossal and Singer, 1968; Valentine et al., 1969).

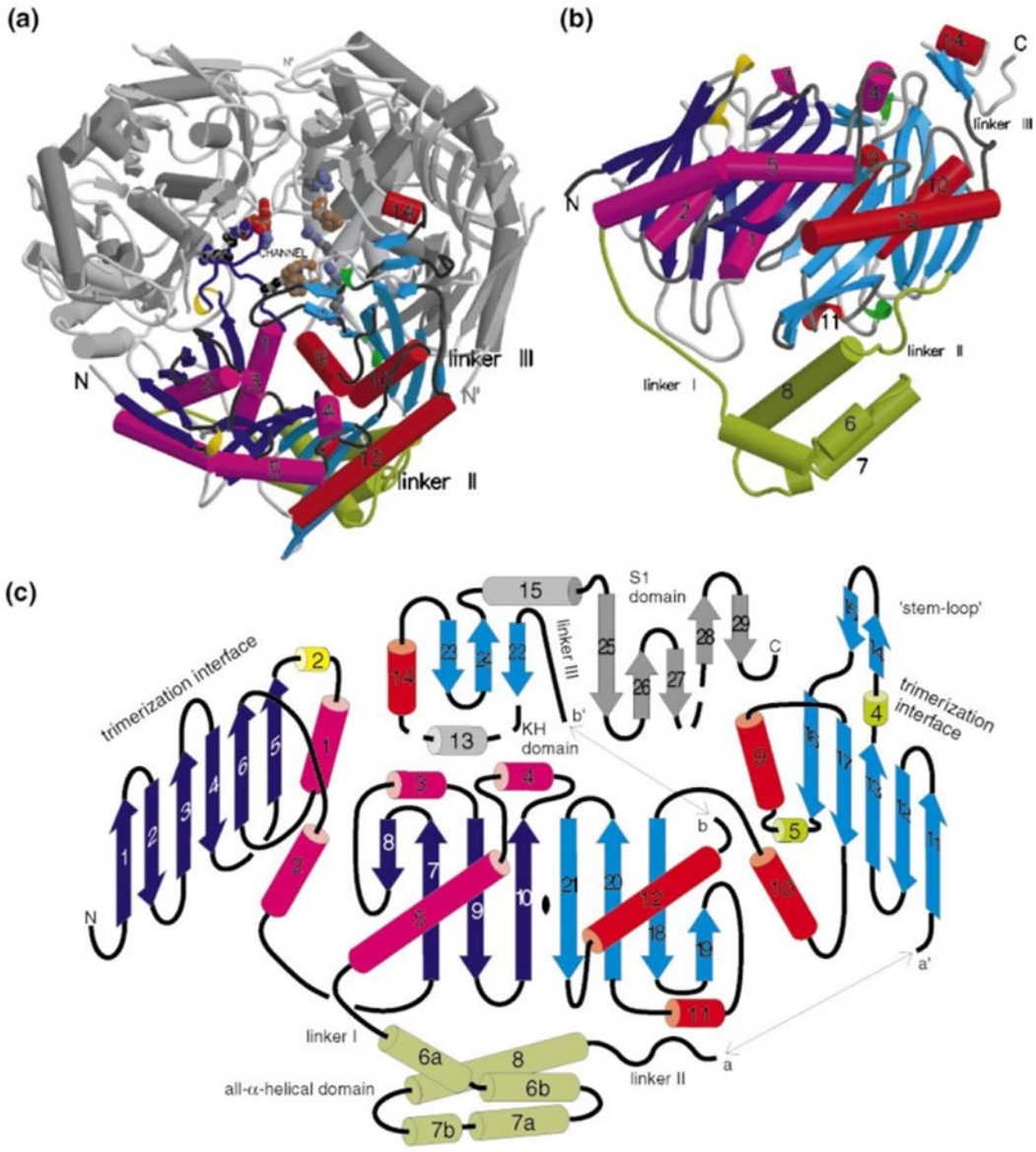

Figure 2 *– The quarternary structure of the PNP enzyme. A) Single subunit in the context of the whole structure. Colour code : red – alpha helix, dark blue – beta sheets, yellow – 3_{10} helices, gray – loops. B) Subunit in detail aligned to the crystallographic 3-fold structure. Colour code : red – alpha helix, dark blue – beta sheets, yellow – 3_{10} helices, gray – loops. C) Topological map of the enzyme structure (Symmons et al., 2000).*

In terms of evolution, PNPase is considered to origin from a duplication event just like the sequentially and functionally similar RNAse PH protein which processes tRNA precursors (Leszczyniecka et al., 2004). From this study, as we can see in figure 3, it appears that the core domain 1 in the last universal common ancestor (LUCA) underwent a duplication event which produced a linked core domain 1 and 2, then copied itself and an upstream stop codon

acquisition followed which blocked the transcription of the core domain 1 as well as the 5′ UTR in RNAse PH being likely silenced (Leszczyniecka et al., 2004).

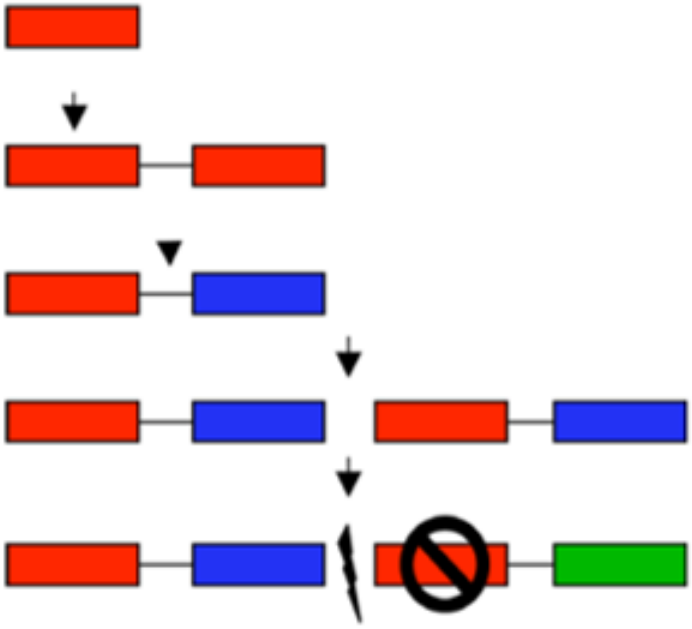

Figure 3 *- Evolution of PNPase via duplication event of core domain 1(red) followed by the production of linked core domain 1 a core domain 2 (blue) which formed the PNPase and RNAse PH (green) after another duplication and finally upstream stop codon. The 5′ UTR in RNAse PH is hypothesized as being silenced (Leszczyniecka et al., 2004).*

Given that PNPase is present in almost every species, including those that lack the other polyadenylation enzyme (PAP) that will be discussed further on, it is likely that PNPase is more ancient that PAP (Slomovic et al., 2006). This can be further supported by research findings (Rott et al., 2003) which confirmed that both cyanobacteria (research done on *Synechocystis*), considered to be a close relative of the bacterial ancestor of the chloroplast and chloroplasts themselves lack PAP and use solely PNPase to polyadenylate their mRNAs, unlike *E.coli* despite both being bacterial cell types. Moreover, the S1 domain of PNPase is noted to be structurally related to bacterial cold shock proteins and both are considered to be derived from an ancestral DNA/RNA-binding protein (Bycroft et al., 1997).

1.2 Poly (A) polymerase

Poly (A) polymerase (PAP) (Edmonds and Abrams, 1960) is the main polyadenylating enzyme in most organisms (O´Hara et al., 1995; Mohanty and Kushner 2000) and also the most well researched from *Escherichia coli* (Rorbach et al., 2014). Unlike the PNPase, PAPase synthesizes almost exclusively homopolymeric tails of adenosine (Yehudai-Resheff et al., 2001) by performing a hydrolysis of ATP into AMP and pyrophosphate (Figure 4) (Mohanty and Kushner, 2010). However not all organisms (*Synechocystis spp.*) or semiautonomic organelles (spinach chloroplast) have the PAP (Rott et al., 2003). In organisms that have both PAP and PNP; PNP has rather a degradatory role with minimal polyadenylation activity which is almost exclusively done by PAP (O´Hara et al., 1995; Mohanty and Kushner, 1999) with the level of PNPase transcripts half-life being increased from 1 minute to 1.5 minutes along with the 30 fold increase in poly (A) levels in *E.coli* (Mohanty and Kushner, 1999).

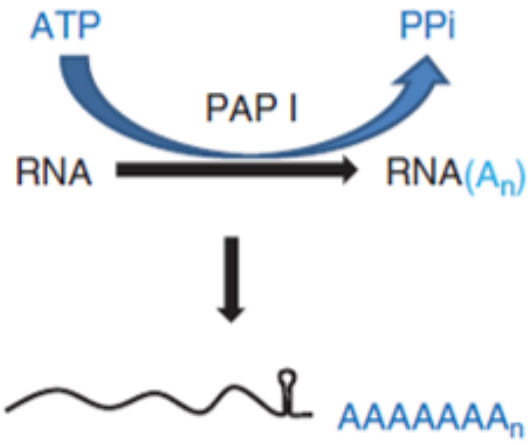

Figure 4 *- PAP mechanism of function – PAP uses ATP hydrolysis to AMP and pyrophosphate (PPi) to synthesize long 3´ poly (A) tails on pre-mRNAs– (Mohanty and Kushner, 2010)*

In terms of its structure, eubacterial PAP composes of an N-terminal domain that contains the catalytic site (approx. 150 amino acids overall), C-terminal RNA binding site and also binding sites for other protein (RNAse E, SrmB) binding in the middle of the sequence. The catalytic site is further split into 3 motifs (M1, M2, M3). Unlike eukaryotic PAP, eubacterial PAP does not need a recognition sequence on the template. Overall, the enzyme consists of 473 amino acids in *E.coli* (Figure 5) (Raynal and Carpousis, 1999). High degree of conservation in terms of structure, as will be discussed further on, was found in the N-terminal

catalytic domain of the PAP enzyme in comparison with tRNA nucleotidyl transferase - ones concerned here repair tRNA by adding a single non-templated adenosine to the CC end, whereas the C-terminal domain tends to differ between organisms (Rott et al., 2003; Masters et al., 1990).

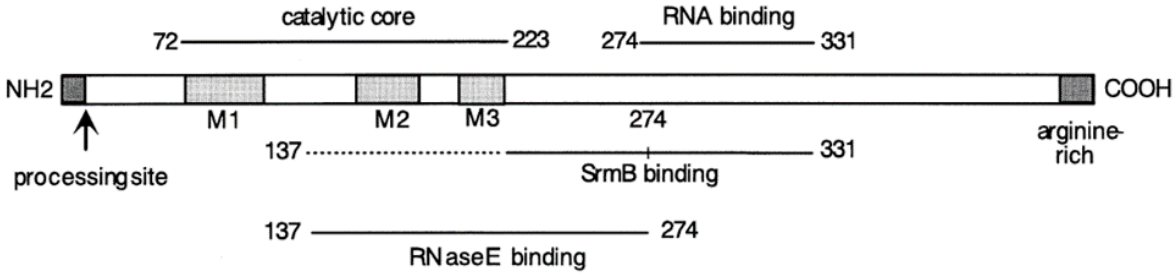

Figure 5 *– Primary structure diagram of the E.coli PAP enzyme. Labelled are the N-terminal processing site, the catalytic site split into motives M1, M2 and M3, the RNA and SrmB protein binding site near the arginine rich C-terminal sequence (Raynal and Carpousis, 1999).*

As to evolution and phylogenetics, a sequence in *Synechocystis spp*. encoding tRNA nucleotidyl transferase that shows high homology to the PAP enzyme was identified (Rott et al in 2003). This tRNA nucleotidyl transferase was observed to add non-templated adenosine to repair the 3´ termini of CCA sequences on the *Synechocystis* tRNA (Rott et al., 2003; Tomita and Yamashita., 2014). *Synechocystis spp.* (Cyanobacteria) as well as the chloroplast organelle lacking degradosome (a complex of RNAse E, PNPase, RNA helicase, enolase and some other RNAs or proteins) and PAP, suggesting that PAP is evolutionary younger than PNP (Rott et al., 2003) and likely evolved from the tRNA nucleotidyl transferase enzyme. Both PNPase and even more significantly PAPase, have been observed to be capable of partially replacing the non-templated adenosine addition repair activity of tRNA nucleotidyl transferase in tRNA nucleotidyl transferase deficient strands of *E.coli*. It has been shown that the time needed for a wild type *E.coli* to double was 27 minutes, tRNA nucleotidyl transferase deficient strain with both enzymes (PAP, PNP) active took 42 minutes, deficient in PAP and tRNA nucleotidyl transferase with functional PNP 70 minutes and deficient in tRNA nucleotidyl transferase and PNPase with functional PAP doubled in 55 minutes. Overexpression of PAP in tRNA nucleotidyl transferase deficient strain increased the growth speed while the overexpression of PNPase in the same setting did not affect the growth speed (Reuven et al., 1997). Overexpression of *E.coli* PAP from a recombinant plasmid also resulted in undesired polyadenylation of the 3´ termini of 23S rRNAs and ultimately cell death. After the induction

of overexpression of the PAP and subsequent 30x increase in poly (A) tails, mRNA half-life decreased from 7.5 minutes in wild-type to 4 minutes. (Mohanty and Kushner, 2012).

Just et al., (2008), further identified a flexible loop structure consisting of 10-12 amino acids on the tRNA nucleotidyl transferase in *E.coli* responsible for adding the non-templated 3′ adenosine which appears to be present in the PAP enzyme as well (even at similar position on the N-terminal), showing that the non-templated adenosine adding structure is indeed present as a rudiment in an evolutionary younger PAP. The subunits that add cytosine and adenosine are two separate entities (Li et al., 2000; Li et al., 2002).

1.3 Taq polymerase

Taq polymerase is a DNA polymerase I from an extremophilic eubacterium *Thermus aquaticus* (Chien et al., 1976). Due to its stability in high temperature, it has been used as the main polymerising enzyme in the Polymerase Chain Reaction (Saiki et al., 1988). However besides classical polymerase activities such as the synthesis of a complementary strand to a DNA template and others, it is also known to add a non-templated adenosine to the 3′ end of the product (Clark, 1988). This ability of the polymerase either viewed as problematic for sequencing (Brownstein et al., 1996) or advantageous in terms of T/A cloning strategy (Holton and Graham, 1991). Nevertheless the reason behind this activity is yet largely unknown.

Firstly, Taq polymerase is certainly not the only polymerase with this activity (Hu, 1993; Garcia et al., 2004; Fiala et al., 2007). Non-templated adenosine addition to the 3′ strand is reported in Vent (adds preferably adenosine), Sequenase (incorporates the same nucleotide as the last template one), *E.coli* DNA polymerase I and its Klenow fragment (weak in comparison to others - less than 10% of the 3′ ends have a non-templated extension). Bacteriophage T4, T7 DNA polymerase and Pfu was found not to have this activity at all. Yet still, Taq polymerase has the strongest non-templated adenosine addition activity out of all of the polymerases mentioned above (Hu, 1993). Hu, (1993) also suggests that the non-templated addition activity might be linked to the lack of 3′ -5′ exonuclease activity since this activity would logically act against the extension of 3′ end.

A research (Brownstein et al., 1996), then sheds light on how is this non-templated 3′ adenosine addition influenced by the structure of the primer in the case of Taq polymerase. According to this work, it appears that Taq polymerase prefers the addition of adenosine next to a 3′ cytosine base while avoiding such addition next to a 3′ adenosine or guanosine. 5′-GTTTCT-3′ reverse primer was found to promote polyadenylation the strongest; resulting in

nearly 100% polyadenylation of all products. The preference of Taq polymerase for non-templated addition of adenosine next to a cytosine base was confirmed in an earlier study along with Sequenase having the same property. Vent polymerase however prefers to add the non-templated base next to a guanosine (Hu, 1993). Also Taq polymerase likely interacts with 6 and more bases on a sequence when adding a 3′ non-templated adenosine (Brownstein et al., 1996).

Considering the mechanism of a non-templated nucleotide addition, it was determined on archea *Sulfolobus solfataricus* that adenine of an arriving dATP stacks against the 5′ strand complementary base which facilitates the synthesis of the 3′ stand and promotes non-templated addition. dATPs are a preferred substrate as they have a strong intramolecular stacking ability (Fiala et al., 2007).

As for the sequence of Taq polymerase, it is 51% homological in its polymerase domain to *E.coli* DNA polymerase I, bearing 6 highly conserved amino acid residues throughout the DNA polymerase family A. The Taq polymerase is composed of 3 domains (polymerase domain, N-terminal 5′ -3′ exonuclease domain and C- terminal 3′ -5′ exonuclease domain), approx. 130 Å long with unusually long shape. These two DNA polymerases are almost identical, although major difference exists between Klenow fragment of *E.coli* DNA polymerase I 3′ -5′ exonuclease domain and Taq pol 3′ -5′ exonuclease domain where Taq polymerase has deletions of four loops of 8-27 residues as well as replacement of catalytic ion-binding carboxylates for dysfunctional ones (Figure 6). Also Taq polymerase has more intermolecular forces in its structure which may explain its higher thermal stability (Kim et al., 1995).

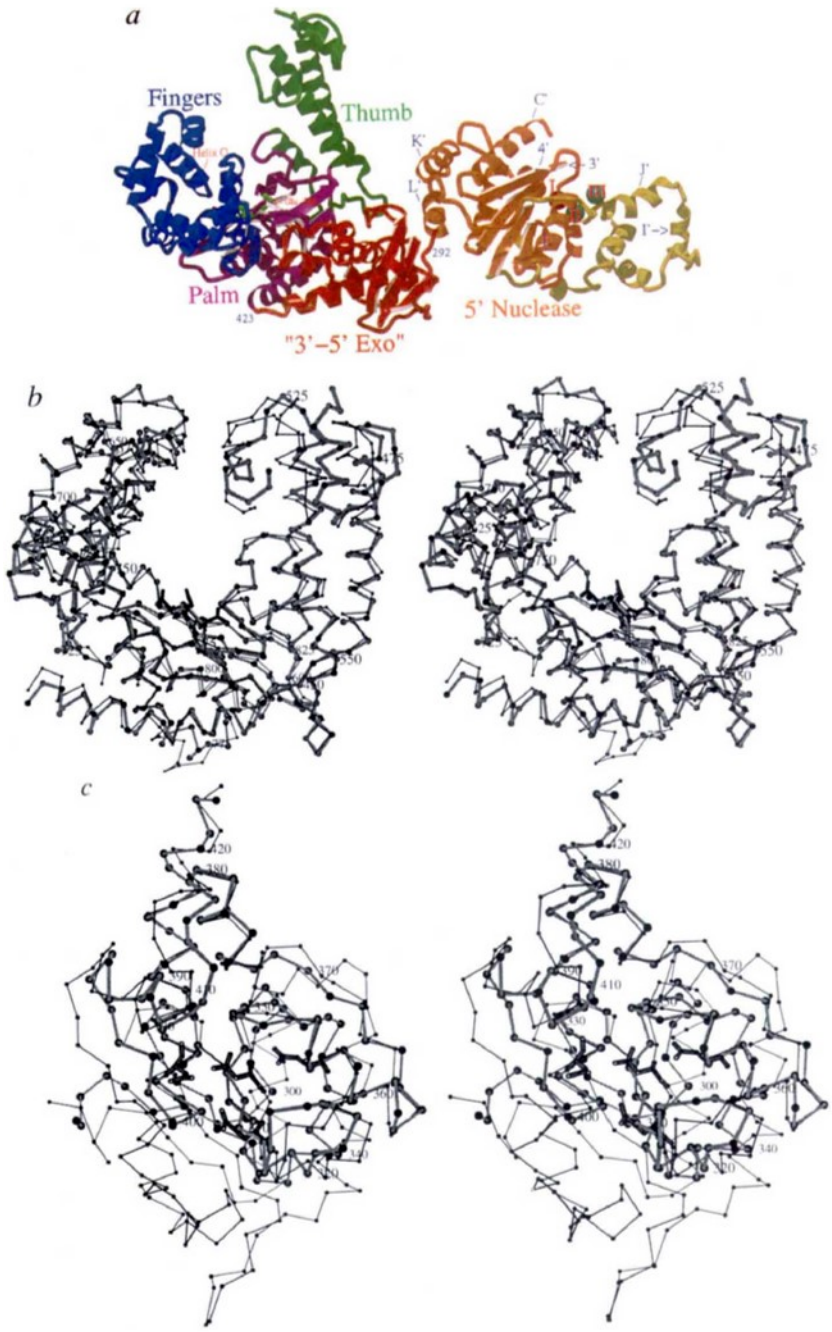

Figure 6 *- Structure of Taq polymerase compared to the structure of Klenow fragment of DNA polymerase I A) Taq polymerase structure diagram – red: 3′ -5′ exo (non-functional), pink: palm domain, blue: fingers domain, light green: thumb domain, yellow and orange: 5′ -3′ exo domain, dark green: polymerase catalytic site; B) Line-up of the Klenow fragment of DNA polymerase I and Taq polymerase domain structures. Thin lines – Klenow fragment, thick lines – Taq polymerase; C) Differing structure of the 3′-5′ exonuclease domain of Klenow fragment (thin lines) and Taq polymerase (thick lines) (Kim et al., 1995).*

Overall, DNA polymerases have been divided into several families based on their amino acid homologies – A family (similar to *E.coli* DNA polymerase I), B family (similar to eukaryotic DNA polymerase alpha and *E.coli* DNA polymerase II), C family (similar to *E.coli* DNA polymerase III) (Ito and Braithwaite, 1991), D family (Ishino et al., 1998), X family (similar to eukaryotic DNA polymerase beta) (Ito and Braithwaite, 1991), and Y family (similar to RAD30 and UmuC gene products) (Filée et al., 2002). Taq polymerase belongs to family A polymerases (Ito and Braithwaite, 1991). PAP belongs to family X polymerases (Balbo and Bohm, 2007) as well as CCA-adding nucleotidyl transferase (Yue et al., 1996).

2 Non-templated 3′ addition of nucleotides in viruses

2.1 Polyadenylation in viruses

Polyadenylation and the non-templated addition of adenosine is not exclusive to eubacteria, archaea and eukaryotes. Research has shown that viruses too are capable of such activities, particularly negative single strand RNA viruses containing RNA dependent RNA polymerase (RdRp) such as Paramyxoviridae (Hausmann et al., 1999; Horikami et al., 1992), Orthomyxoviridae (Zheng et al., 1999), Rhabdoviridae (Schneemann et al., 1994) and Filoviridae (Mühlberger, 2007) with a family of positive single strand RNA viruses (Potyviridae) recently also described to have this activity (Olspert et al., 2015; Rodamilans et al., 2015).

These viruses appear to have a special sequence of usually 5-7 uridines at the 5′ ends of the negative single strand RNA genome (but can be up to 13 nucleotides) which, once encountered by the RdRp, causes the enzyme to jam in place and scan the same sequence over and over, adding the same nucleotides (Zheng et al., 1999), producing a 3′ poly (A) chain around 100-300 nucleotides long on the complementary strand (Figure 7) (Hausmann et al., 1999; Whelan et al., 2004). This phenomenon is labelled as polymerase stuttering (Vidal et al., 1990) or Markov model of polymerase stuttering (Jadhav et al., 2020) and may also happen in a region containing cytosine residues (Zheng et al., 1999). Although subject to discussion, the polymerase stuttering can be considered a non-templated addition of a nucleotides, as the polymerase does not exactly copy the template. It is also important to note that the uridine rich sequence is always preceded by a cytosine nucleotide, often part of a conserved sequence 3′ -AUAC- 5′. The cytosine appears to have a key role in the termination of transcription as when changes are made to the particular nucleotide in vesicular stomatitis virus, the termination signal is neutralized which isn′t the case with changes to other nucleotides presents in the sequence mentioned (Whelan et al., 2004; Barr et al., 1997). At the same time, the 3′ -AUAC- 5′

sequence in vesicular stomatitis virus genome needs to be placed directly in front of the uridine sequence, otherwise the polymerase will read through the sequence and won´t terminate transcription. Nevertheless, the polyadenylation was found to be necessary but not sufficient by itself to initiate termination (Barr et al., 1997). Another study suggests that apart from the poly (A) formation signal, the sequence of 3´ -UCAAU- 5´ in respiratory syncytial virus genome plays a role and length of a central AU rich region too (Harmon et al., 2001; Sutherland et al., 2001).

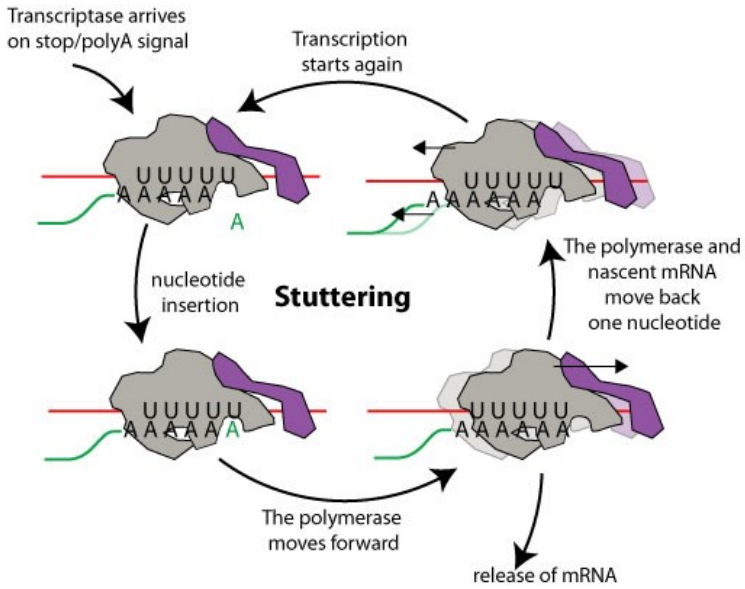

Figure 7 *– Mechanism of polymerase stuttering. The polymerase reaches the slippery sequence, adds a nucleotide, moves one nucleotide forward on template, dissociates from the template and moves back one nucleotide back, adding the same nucleotide (Swiss Institute of Bioinformatics, 2022 - https://viralzone.expasy.org/1916)*

2.2 Co-transcriptional mRNA editing in viruses

The negative strand RNA viruses not only use this polymerase stuttering mechanism to polyadenylate 3´ ends of their mRNAs but also to induce a frameshift as a form of RNA co-transcriptional editing (Whelan et al., 2004; Thomas et al., 1988; Jacques and Kolakofsky, 1991). Due to that, we can see several different products from a single gene that share the N-terminal region yet differ in the C-terminal region which is the result of the polymerase stuttering and adding a different number of non-templated guanosines, in some cases even 7 or more (Douglas et al., 2021; Thomas et al., 1988; Jacques and Kolakofsky,1991).

Overall, there are slight differences between individual negative single strand RNA viruses, such as the editing site sequence structure (Jacques et al., 1994) yet the main principle

remains the same. For example, Paramyxoviridae members, like other negative single strand RNA viruses, edit their P gene mRNA to code for several different proteins from the same sequence via expansion of the number of G´s in a conserved A_nG_n transcript sequence which shifts the reading frame. The number of G´s inserted into the transcript in this case depends on the viral species and circumstances – sendai virus naturally tends to insert only a single nucleotide while mumps virus inserts two and parainfluenza type 3 inserts up to six nucleotides (Hausmann et al., 1999).

Regarding the frequency of editing, wild type-mRNAs (A_6G_3) have 30-50% of the mRNAs edited in sendai virus (Vidal et al., 1990; Jacques et al., 1994). When the number of adenosines at the editing site of a sendai virus transcript is experimentally reduced (A_5G_4 or A_3G_6), the editing becomes low (7% of mRNA edited) or inactive, perhaps due to the formation of unstable A/C pairing. On the contrary, increasing the amount of adenosines in the sequence (A_7G_2) increased the editing frequency to 67% of all mRNAs being edited and 12% of them had multiple guanosines added, even up to 20 nucleotides in the case of A_8G_1. A_9G_0 does not appear to be viable. Overall, mutating guanosines into adenosines in the run increases the editing frequency (up to 67%) yet mutating adenosines into guanosines has the opposite effect (Hausmann et al., 1999).

It was also found that upstream cis-acting sequences direct if and how many times does the RdRp stutter on the editing site during the mRNA co-transcriptional editing. $^{-11}$AUU^{-9} increases insertion of a single G in comparison to wild type $^{-11}$AAC^{-9} up to 53%, not much multi G editing is seen – max. 9%, just like in the case of the poly (A) synthesis. Position closest to the editing sequence is the most important regarding this phenomenon, likely due to the sequence being close to the exit channel of the polymerase when it reads the editing site, thus having steric influence (Hausmann et al., 1999). Tampering with these sequences renders the virus (sendai virus, measles virus) unable to perform RNA editing (Schneider et al., 1997; Jacques et al., 1994). Nucleocapsid protein subunits in sendai virus appear to influence the editing as well (Iseni et al., 2002). The editing sequence itself then also controls some of the single G insertions (Hausmann et al., 1999).

Co-transcriptional RNA editing occurs in bacteria, namely in DNA dependent RNA polymerase of *Thermus thermophilus* poly (A) run in its mRNA and *E.coli* on a T_5C_5 slippery sequence too (Larsen et al., 2000, Penno et al., 2015). And interestingly, Klenow fragment of DNA polymerase I was observed to stutter at poly (G) runs in the P protein coding sequence – and when Sequenase was used along with dITPs instead of dGTPs, no stuttering was observed.

However if the slippery sequence had a strong secondary structure, the Sequenase became even more prone to stuttering than the Klenow fragment of DNA polymerase I (Thomas et al., 1988).

When the guanosine triphosphate is partially replaced by inositol triphosphate in combination with natural RNA dependent RNA polymerase, the amount of edited P genes in sendai virus rises significantly; up to 10 fold (Vidal et al., 1990). Why does this occur is unknown currently; might be due to the different chemical properties of inositol opposed to guanosine. But it does indicate that there are clear differences in response to environmental conditions between non-templated activity prone polymerases such as Sequenase, DNA polymerase I Klenow fragment and RNA dependent RNA polymerase.

Next, as proven by research (Zheng et al., 1999), RdRp stuttering isn´t unique only to transcription of negative single strand RNA viruses. It can unintentionally occur during replication as well if the RdRp doesn´t ignore the termination signal or co-transcriptional edit site as it usually does but the exact reason why this happens isn´t entirely understood yet (Douglas et al., 2021).

2.3 Negative single strand RNA virus RNA dependent RNA polymerase

RNA dependent RNA polymerase is an enzyme almost unique to RNA viruses which code this polymerase to be able to replicate in host cells (te Velthuis et al., 2021).

Focusing on structure, up until recently, not many studies have been done on the negative single strand RNA virus RdRp structure (Fearns and Plemper, 2017) as the process of determining the structure of RdRp was very challenging however that is starting to change in the recent years with improving microscopy technology (te Velthuis et al., 2021), therefore the older sources usually integrate fragments of knowledge regarding different viral species.

The typical negative single strand RNA virus RdRp is approximately 250 kDa, contains about 2000 amino acids and has two separate catalytic modes – replication and transcription (Fearns and Plemper, 2017).

The structure of the non-segmented negative single strand RNA virus RdRp can be divided into a large conserved subunit serving as a catalytic domain (L) and a diverse phosphoprotein subunit serving as co-factor (P) (Horikami et al., 1992). The L subunit of vesicular stomatitis virus polymerase appears to be of a classical right hand shape with palm, thumb and finger subdomains, containing two channels for nucleotide and template entry (Figure 8) (Liang et al., 2015). N-terminal region of the L protein appears to interact with the P

subunit and is less conserved than the rest of the L protein for the measles virus (Horikami et al., 1994).

Influenza A virus RdRp, one of the segmented viruses, for instance, consists of a heterotrimer labelled as PB1, PB 2 and PA (te Velthuis et al., 2021; Zheng et al., 1999). The PB1 is the catalytic centre while the PA subunit is then divided into N-terminal endonuclease domain and C-terminal domain which forms the thumb (te Velthuis et al., 2021). RdRp and endonuclease domain are the most conserved parts of the polymerase in the viral order (Olschewski et al., 2020). The active site of the RdRp domain is structurally almost identical in all of the viral RdRp yet not always sequentially (Gogrefe et al., 2019).

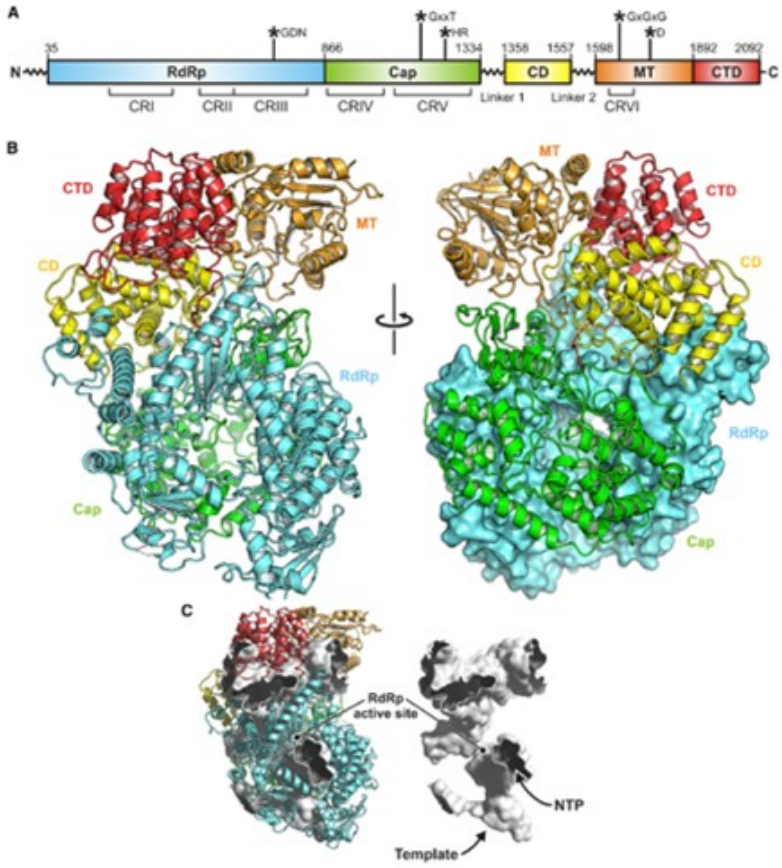

Figure 8 *- Structure of vesicular stomatitis virus RNA polymerase. Cyan – catalytic site, green - 5′ cap synthesis site, yellow – connecting domain, orange – methyltransferase domain, red – C-terminal domain. Diagram A) overview of the domain organisation (CR I-IV – conserved regions), Diagram B) structure diagram, Diagram C) substrate channel depiction diagram (Liang et al., 2015).*

The evolutionary history of RdRp is likely connected to the evolutionary history of viruses themselves which have been a phylogenetic challenge for a considerable amount of time

(Holmes, 2011). The main reason why it has been challenging to establish evolutionary relationships between viruses is their high mutation rate (Suttle, 2005). Having high mutation rate has an evolutionary advantage; it enables the modern viruses to specialise quickly and escape host defence mechanisms. RdRp is ideal for that purpose as it is prone to errors (Vignuzzi et al., 2006; Holmes, 2011) and lacks proofreading mechanisms, as research on vesicular stomatitis virus RdRp shows (Steinhauer et al., 1992).

Many studies have noted that although quite diverse, all polymerases (RNA and DNA) share a basic highly conserved core in the palm subdomain that forms the catalytic unit of these enzymes (Delarue et al., 1990; Joyce and Steitz, 1994; Hansen et al., 1997). This unit is formed by an aspartate sequence which interacts with Mg^{2+} ions and that allows the process of nucleotide addition to occur (de Farias et al., 2017; Jeruzalmi and Steitz, 1998).

Given the existence of this conserved core in polymerases all across the life domains (Delarue et al., 1990; Joyce and Steitz, 1994; Hansen et al., 1997) and the almost exclusive presence of RdRp in RNA viruses (te Velthuis et al., 2021), the possibility of RNA dependent RNA polymerase being ancestral to all other polymerases, or at least the modern RdRp being the closest relative of the universal ancestral polymerase becomes logically prominent.

This is further supported by research (de Farias et al., 2016b) who proposes the ancient proteome before LUCA was based on ancestral tRNA translation, amazingly discovered that after comparisons of modern proteins and ancient tRNA sequences, the RdRp was by far the most related. This finding has quite an implication for the evolution of viruses, proteosynthetic system and the central topic of this work and will be discussed further in the discussion section.

Discussion and conclusion

The non-templated addition of nucleotide, once regarded as a sole nuisance during genotyping (Brownstein et al., 1996) appears to be a very interesting phenomenon involving the phylogenetics of cellular life and viruses.

Firstly, the two polyadenylation enzymes present in bacteria, the PNPase and the PAPase, appear to have differing origins, with PNPase core 1 being a result of a duplication event on the level of LUCA. It is noted the core being similar to RNAse PH which processes tRNA precursors (Leszczyniecka et al., 2004). The second similarity in the structure of PNPase is of the S1 unit that is similar to bacterial cold shock proteins that arose from ancestral DNA/RNA binding proteins (Bycroft et al., 1997).

PAPase N- terminal domain containing the polymerase catalytic site, on the other hand, has been shown to have descended much more recently from CCA tRNA nucleotidyl transferase, an enzyme that adds a non-templated adenosine to the CC sequence of tRNA (again an enzyme involved in tRNA processing) and is still able to partially perform its progenitor´s function (Reuven et al., 1997). The sequence responsible for the addition of the adenosine was identified in both of the enzymes (Just et al., 2008).

Research done on Taq polymerase and other polymerases commonly used in molecular biology (Sequenase, Vent, Klenow fragment, T4, Pfu) suggests that the tendency to add non-templated adenosine is dependent on the strength of the 3´ -5´ exonuclease activity of the enzymes concerned, given the enzymes most prone to this behaviour have a weak or absent 3´ -5´ exonuclease activity. RNA dependent RNA polymerases are said to lack 3´ -5´ exonuclease activity as a whole (Steinhauer et al., 1992) and also prone to non-templated nucleotide additions (Vidal et al., 1990). A difference in preference of adjacent nucleotides was noted – the bacterial (Taq polymerase) and bacteriophage (Sequenase) polymerases prefer to add adenosine next to a cytosine base while the archaeal Vent polymerase prefers to add adenosine adjacent to guanosine base (Hu, 1993). Similar occurrence can be observed in the RdRp stuttering polyadenylation where the template short uridine run is always preceded by a cytosine (Whelan et al., 2004; Barr et al., 1997). The significance of this observation or whether it has any implications on phylogeny is uncertain as of now. In both the mentioned DNA polymerases and viral RdRp, the enzymes have been suggested to interact with more than one base at a time which influences their function (Hausmann et al., 1999; Brownstein et al., 1996). In viral RdRp, co-transcriptional editing and polyadenylation by stuttering have been suggested to be evolutionary linked (Hausmann et al., 1999). Another major influence on the stuttering behaviour of these polymerases is the strength of the template secondary structure (Thomas et al., 1988) and by the nature of the nucleotides in the environment (Vidal et al., 1990).

Given all these similarities in the non-templated nucleotide adding behaviour of the polymerases, for example the ability of Sequenase and Klenow fragment of DNA polymerase I to stutter on RdRp stuttering sites given the right conditions (Thomas et al., 1988) and the fact that all polymerases contain a conserved core of the catalytic site (Delarue et al., 1990; Joyce and Steitz, 1994; Hansen et al., 1997), it appears to me that the tendency to incorporate non-templated nucleotides under given circumstances is somehow encoded in the conserved structures of the polymerases and therefore preserved in all of them. What differs is of course the way this property is handled – some polymerases evolved mechanisms (strong 3´-5´

exonuclease activity) to counteract the tendency while others took advantage of it like RNA-dependent RNA polymerase in negative single strand RNA viruses as fast mutation is important for their survival (Vignuzzi et al., 2006; Holmes, 2011).

Therefore the ability to add non-templated nucleotide of Taq and certain other polymerases appears to be a rudimentary function yet not a rudiment of ancient polyadenylation activity. It rather is a function inherited from an RNA dependent RNA polymerase – like ancestor that was concealed by the formation of 3´-5´ exonuclease activity and exposed when Taq polymerase acquired a deletion in the sequence for the 3´-5´ exonuclease activity (Kim et al., 1995).

In my opinion, these findings can be further used in researching the origin of the proteosynthetic apparatus and RNA viruses, as it appears to support the theory of RNA preceding the emergence of DNA as a memory molecule, given that it was discovered to have catalytic functions as well (Kruger et al., 1982). According to research (de Farias et al., 2017) RdRp originates from junctions of ancestral tRNAs that based on this author formed the first ever genes – therefore from the time of a pre-LUCA world. Given that RNA viruses and viruses in general are so genetically distinct (Suttle, 2005) from the rest of the biological systems and lateral gene transfer with their host, although occurring on occasion, is rather rare (Holmes, 2009), it is likely that their origin precedes the LUCA. What properties could a proto-RNA virus have (besides the RNA dependent RNA polymerase) is a matter of speculation. Proto-RNA viruses at the time of their emergence didn´t necessarily have to occupy their current niche as an obligate parasite – after all, there isn´t anything to indicate that these viruses never had translation apparatus. They simply could have lost it secondarily after adapting to the role of a cellular parasite. Such way of thinking offers the question whether the specialisation of proto-RNA viruses into cellular parasites didn´t trigger the transition of non-parasitic proto-cells to DNA which could offer a survival advantage. Certainly more research needs to be done to answer these crucial questions.

Literature

Anantharaman, V. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Research*, *30*(7), 1427–1464. https://doi.org/10.1093/nar/30.7.1427

Balbo, P. B., & Bohm, A. (2007). Mechanism of poly(a) polymerase: Structure of the enzyme-mgatp-RNA ternary complex and kinetic analysis. *Structure*, *15*(9), 1117–1131. https://doi.org/10.1016/j.str.2007.07.010

Barr, J. N., Whelan, S. P., & Wertz, G. W. (1997). Cis-acting signals involved in termination of vesicular stomatitis virus mrna synthesis include the conserved AUAC and the U7 signal for polyadenylation. *Journal of Virology*, *71*(11), 8718–8725. https://doi.org/10.1128/jvi.71.11.8718-8725.1997

Brownstein, M. J., Carpten, J. D., & Smith, J. R. (1996). Modulation of non-templated nucleotide addition by *taq* DNA polymerase: Primer modifications that facilitate genotyping. *BioTechniques*, *20*(6), 1004–1010. https://doi.org/10.2144/96206st01

Bycroft, M., Hubbard, T. J. P., Proctor, M., Freund, S. M. V., & Murzin, A. G. (1997). The solution structure of the S1 RNA binding domain: A member of an ancient nucleic acid–binding fold. *Cell*, *88*(2), 235–242. https://doi.org/10.1016/s0092-8674(00)81844-9

Charlesworth, A., Meijer, H. A., & de Moor, C. H. (2013). Specificity factors in cytoplasmic polyadenylation. *Wiley Interdisciplinary Reviews: RNA*, *4*(4), 437–461. https://doi.org/10.1002/wrna.1171

Chien, A., Edgar, D. B., & Trela, J. M. (1976). Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus. *Journal of Bacteriology*, *127*(3), 1550–1557. https://doi.org/10.1128/jb.127.3.1550-1557.1976

Clark, J. M. (1988). Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Research*, *16*(20), 9677–9686. https://doi.org/10.1093/nar/16.20.9677

Colgan, D. F., & Manley, J. L. (1997). Mechanism and regulation of mrna polyadenylation. *Genes & Development*, *11*(21), 2755–2766. https://doi.org/10.1101/gad.11.21.2755

Cui, J., Sartain, C. V., Pleiss, J. A., & Wolfner, M. F. (2013). Cytoplasmic polyadenylation is a major mrna regulator during oogenesis and egg activation in drosophila. *Developmental Biology*, *383*(1), 121–131. https://doi.org/10.1016/j.ydbio.2013.08.013

de Farias, S. T., Rêgo, T. G., & José, M. V. (2016). A proposal of the proteome before the last universal common ancestor (luca). *International Journal of Astrobiology*, *15*(1), 27–31. https://doi.org/10.1017/s1473550415000464

de Farias, S. T., dos Santos Junior, A. P., Rêgo, T. G., & José, M. V. (2017). Origin and evolution of RNA-dependent RNA polymerase. *Frontiers in Genetics*, *8*. https://doi.org/10.3389/fgene.2017.00125

Delarue, M., Poch, O., Tordo, N., Moras, D., & Argos, P. (1990). An attempt to unify the structure of polymerases. *"Protein Engineering, Design and Selection"*, *3*(6), 461–467. https://doi.org/10.1093/protein/3.6.461

Douglas, J., Drummond, A. J., & Kingston, R. L. (2021). Evolutionary history of cotranscriptional editing in the paramyxoviral phosphoprotein gene. *Virus Evolution*, *7*(1). https://doi.org/10.1093/ve/veab028

Edmonds, M., & Abrams, R. (1960). Polynucleotide biosynthesis: Formation of a sequence of adenylate units from adenosine triphosphate by an enzyme from thymus nuclei. *Journal of Biological Chemistry*, *235*(4), 1142–1149. https://doi.org/10.1016/s0021-9258(18)69494-3

Fearns, R., & Plemper, R. K. (2017). Polymerases of paramyxoviruses and pneumoviruses. *Virus Research*, *234*, 87–102. https://doi.org/10.1016/j.virusres.2017.01.008

Fiala, K. A., Brown, J. A., Ling, H., Kshetry, A. K., Zhang, J., Taylor, J.-S., Yang, W., & Suo, Z. (2007). Mechanism of template-independent nucleotide incorporation catalyzed by a template-dependent DNA polymerase. *Journal of Molecular Biology*, *365*(3), 590–602. https://doi.org/10.1016/j.jmb.2006.10.008

Filée, J., Forterre, P., Sen-Lin, T., & Laurent, J. (2002). Evolution of DNA polymerase families: Evidences for multiple gene exchange between cellular and viral proteins. *Journal of Molecular Evolution*, *54*(6), 763–773. https://doi.org/10.1007/s00239-001-0078-x

Garcia-Mena, J., Das, A., Sanchez-Trujillo, A., Portier, C., & Montanez, C. (1999). A novel mutation in the KH domain of polynucleotide phosphorylase affects autoregulation and mrna decay in escherichia coli. *Molecular Microbiology*, *33*(2), 235–248. https://doi.org/10.1046/j.1365-2958.1999.01451.x

García, P. B., Robledo, N. L., & Islas, Á. L. (2004). Analysis of non-template-directed nucleotide addition and template switching by DNA polymerase. *Biochemistry*, *43*(51), 16515–16524. https://doi.org/10.1021/bi0491853

Gibson, T. J., Thompson, J. D., & Heringa, J. (1993). The kh domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NUSA and is probably involved in binding to nucleic acid. *FEBS Letters*, *324*(3), 361–366. https://doi.org/10.1016/0014-5793(93)80152-k

Gogrefe, N., Reindl, S., Gunther, S., & Rosenthal, M. (2019). Structure of the cap-binding domain of Rift Valley fever virus L protein. https://doi.org/10.2210/pdb6qhg/pdb

Grunberg-Manago, M., Ortiz, P. J., & Ochoa, S. (1955). Enzymatic synthesis of Nucleic Acidlike polynucleotides. *Science*, *122*(3176), 907–910. https://doi.org/10.1126/science.122.3176.907

Hansen, J. L., Long, A. M., & Schultz, S. C. (1997). Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure*, *5*(8), 1109–1122. https://doi.org/10.1016/s0969-2126(97)00261-x

Harmon, S. B., Megaw, A. G., & Wertz, G. W. (2001). RNA sequences involved in transcriptional termination of respiratory syncytial virus. *Journal of Virology*, *75*(1), 36–44. https://doi.org/10.1128/jvi.75.1.36-44.2001

Hausmann Stéphane, Garcin, D., Delenda, C., & Kolakofsky, D. (1999). The versatility of paramyxovirus RNA polymerase stuttering. *Journal of Virology*, *73*(7), 5568–5576. https://doi.org/10.1128/jvi.73.7.5568-5576.1999

Holmes, E. C. (2009). The evolution and emergence of RNA viruses. *Emerging Infectious Diseases*, *16*(5), 899–899. https://doi.org/10.3201/eid1605.100164

Holmes, E. C. (2011). What does virus evolution tell us about virus origins? *Journal of Virology*, *85*(11), 5247–5251. https://doi.org/10.1128/jvi.02203-10

Holton, T. A., & Graham, M. W. (1991). A simple and efficient method for direct cloning of PCR products using DDT-tailed vectors. *Nucleic Acids Research*, *19*(5), 1156–1156. https://doi.org/10.1093/nar/19.5.1156

Horikami, S. M., Curran, J., Kolakofsky, D., & Moyer, S. A. (1992). Complexes of sendai virus NP-P and P-L proteins are required for defective interfering particle genome replication in vitro. *Journal of Virology*, *66*(8), 4901–4908. https://doi.org/10.1128/jvi.66.8.4901-4908.1992

Horikami, S. M., Smallwood, S., Bankamp, B., & Moyer, S. A. (1994). An amino-proximal domain of the L protein binds to the P protein in the measles virus RNA polymerase complex. *Virology*, *205*(2), 540–545. https://doi.org/10.1006/viro.1994.1676

Hu, G. (1993). DNA polymerase-catalyzed addition of nontemplated extra nucleotides to the 3′ of a DNA fragment. *DNA and Cell Biology*, *12*(8), 763–770. https://doi.org/10.1089/dna.1993.12.763

Iseni, F., Baudin, F., Garcin, D., Marq, J., Ruigrok, R. W. H., & Kolakofsky, D. (2002). Chemical modification of nucleotide bases and mrna editing depend on hexamer or nucleoprotein phase in Sendai virus nucleocapsids. *RNA*, *8*(8), 1056–1067. https://doi.org/10.1017/s1355838202029977

Ishino, Y., Komori, K., Cann, I. K., & Koga, Y. (1998). A novel DNA polymerase family found in*archaea*. *Journal of Bacteriology*, *180*(8), 2232–2236. https://doi.org/10.1128/jb.180.8.2232-2236.1998

Ito, J., & Braithwaite, D. K. (1991). Compilation and alignment of DNA polymerase sequences. *Nucleic Acids Research*, *19*(15), 4045–4057. https://doi.org/10.1093/nar/19.15.4045

Jacques, J. P., & Kolakofsky, D. (1991). Pseudo-templated transcription in prokaryotic and eukaryotic organisms. *Genes & Development*, *5*(5), 707–713. https://doi.org/10.1101/gad.5.5.707

Jacques, J. P., Hausmann, S., & Kolakofsky, D. (1994). Paramyxovirus mrna editing leads to G deletions as well as insertions. *The EMBO Journal*, *13*(22), 5496–5503. https://doi.org/10.1002/j.1460-2075.1994.tb06884.x

Jadhav, A., Zhao, L., Ledda, A., Liu, W., Ding, C., Nair, V., & Ferretti, L. (2020). Patterns of RNA editing in Newcastle disease virus infections. *Viruses*, *12*(11), 1249. https://doi.org/10.3390/v12111249

Jeruzalmi, D., & Steitz, T. A. (1998). Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *The EMBO Journal*, *17*(14), 4101–4113. https://doi.org/10.1093/emboj/17.14.4101

Jordan, P. C., Liu, C., Raynaud, P., Lo, M. K., Spiropoulou, C. F., Symons, J. A., Beigelman, L., & Deval, J. (2018). Initiation, extension, and termination of RNA synthesis by a paramyxovirus polymerase. *PLOS Pathogens*, *14*(2). https://doi.org/10.1371/journal.ppat.1006889

Joyce, C. M., & Steitz, T. A. (1994). Function and structure relationships in DNA polymerases. *Annual Review of Biochemistry*, *63*(1), 777–822. https://doi.org/10.1146/annurev.bi.63.070194.004021

Just, A., Butter, F., Trenkmann, M., Heitkam, T., Mörl, M., & Betat, H. (2008). A comparative analysis of two conserved motifs in bacterial poly(a) polymerase and CCA-adding enzyme. *Nucleic Acids Research*, *36*(16), 5212–5220. https://doi.org/10.1093/nar/gkn494

Kim, Y., Eom, S. H., Wang, J., Lee, D.-S., Suh, S. W., & Steitz, T. A. (1995). Crystal structure of Thermus aquaticus DNA polymerase. *Nature*, *376*(6541), 612–616. https://doi.org/10.1038/376612a0

Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., & Cech, T. R. (1982). Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, *31*(1), 147–157. https://doi.org/10.1016/0092-8674(82)90414-7

Larsen, B., Wills, N. M., Nelson, C., Atkins, J. F., & Gesteland, R. F. (2000). Nonlinearity in genetic decoding: Homologous DNA replicase genes use alternatives of transcriptional

slippage or translational frameshifting. *Proceedings of the National Academy of Sciences*, *97*(4), 1683–1688. https://doi.org/10.1073/pnas.97.4.1683

Leszczyniecka, M., DeSalle, R., Kang, D.-chul, & Fisher, P. B. (2004). The origin of polynucleotide phosphorylase domains. *Molecular Phylogenetics and Evolution*, *31*(1), 123–130. https://doi.org/10.1016/j.ympev.2003.07.012

Li, F., Wang, J., & Steitz, T. A. (2000). Sulfolobus shibatae CCA-adding enzyme forms a tetramer upon binding two trna molecules: A scrunching-shuttling model of CCA specificity. *Journal of Molecular Biology*, *304*(3), 483–492. https://doi.org/10.1006/jmbi.2000.4189

Li, F., Xiong, Y., Wang, J., Cho, H. D. D., Tomita, K., Weiner, A. M., & Steitz, T. A. (2002). Crystal structures of the bacillus stearothermophilus CCA-adding enzyme and its complexes with ATP or CTP. *Cell*, *111*(6), 815–824. https://doi.org/10.1016/s0092-8674(02)01115-7

Liang, B., Li, Z., Jenni, S., Rameh, A. A., Morin, B. M., Grant, T., Grigorieff, N., Harrison, S. C., & Whelan, S. P. J. (2015). Structure of the L protein of vesicular stomatitis virus from electron cryomicroscopy. https://doi.org/10.2210/pdb5a22/pdb

Liu, J., & Schwartz, J. H. (2003). The cytoplasmic polyadenylation element binding protein and polyadenylation of messenger RNA in Aplysia neurons. *Brain Research*, *959*(1), 68–76. https://doi.org/10.1016/s0006-8993(02)03729-0

Luitjens, C., Gallegos, M., Kraemer, B., Kimble, J., & Wickens, M. (2000). CPEB proteins control two key steps in spermatogenesis in *C. elegans*. *Genes & Development*, *14*(20), 2596–2609. https://doi.org/10.1101/gad.831700

Masters, M., March, J. B., Oliver, I. R., & Collins, J. F. (1990). A possible role for the PCNB gene product of escherichia coli in modulating RNA:RNA interactions. *Molecular and General Genetics MGG*, *220*(2), 341–344. https://doi.org/10.1007/bf00260507

Mohanty, B. K., & Kushner, S. R. (1999). Analysis of the function of escherichia coli poly(a) polymerase I in RNA metabolism. *Molecular Microbiology*, *34*(5), 1094–1108. https://doi.org/10.1046/j.1365-2958.1999.01673.x

Mohanty, B. K., & Kushner, S. R. (2000). Polynucleotide phosphorylase functions both as a 3′ → 5′ exonuclease and a poly(a) polymerase in *escherichia coli*. *Proceedings of the National Academy of Sciences*, *97*(22), 11966–11971. https://doi.org/10.1073/pnas.220295997

Mohanty, B. K., & Kushner, S. R. (2010). Bacterial/archaeal/organellar polyadenylation. *Wiley Interdisciplinary Reviews: RNA*, *2*(2), 256–276. https://doi.org/10.1002/wrna.51

Mohanty, B. K., & Kushner, S. R. (2012). Deregulation of poly(a) polymerase I in escherichia coli inhibits protein synthesis and leads to cell death. *Nucleic Acids Research*, *41*(3), 1757–1766. https://doi.org/10.1093/nar/gks1280

Mühlberger, E. (2007). Filovirus replication and transcription. *Future Virology*, *2*(2), 205–215. https://doi.org/10.2217/17460794.2.2.205

Nossal, N. G., & Singer, M. F. (1968). The processive degradation of individual polyribonucleotide chains. *Journal of Biological Chemistry*, *243*(5), 913–922. https://doi.org/10.1016/s0021-9258(18)93603-3

O'Hara, E. B., Chekanova, J. A., Ingle, C. A., Kushner, Z. R., Peters, E., & Kushner, S. R. (1995). Polyadenylylation helps regulate mrna decay in escherichia coli. *Proceedings of the National Academy of Sciences*, *92*(6), 1807–1811. https://doi.org/10.1073/pnas.92.6.1807

Olschewski, S., Cusack, S., & Rosenthal, M. (2020). The Cap-snatching mechanism of Bunyaviruses. *Trends in Microbiology*, *28*(4), 293–303. https://doi.org/10.1016/j.tim.2019.12.006

Olspert, A., Chung, B. Y. W., Atkins, J. F., Carr, J. P., & Firth, A. E. (2015). Transcriptional slippage in the positive-sense rna virus family *potyviridae*. *EMBO Reports*, *16*(8), 995–1004. https://doi.org/10.15252/embr.201540509

Penno, C., Sharma, V., Coakley, A., O'Connell Motherway, M., van Sinderen, D., Lubkowska, L., Kireeva, M. L., Kashlev, M., Baranov, P. V., & Atkins, J. F. (2015). Productive mrna stem loop-mediated transcriptional slippage: Crucial features in common with intrinsic

terminators. *Proceedings of the National Academy of Sciences*, *112*(16). https://doi.org/10.1073/pnas.1418384112

Portnoy, V., Evguenieva-Hackenberg, E., Klein, F., Walter, P., Lorentzen, E., Klug, G., & Schuster, G. (2005). RNA polyadenylation in archaea: Not observed in *haloferax* while the exosome polynucleotidylates RNA in *sulfolobus*. *EMBO Reports*, *6*(12), 1188–1193. https://doi.org/10.1038/sj.embor.7400571

Portnoy, V., & Schuster, G. (2008). Mycoplasma gallisepticum as the first analyzed bacterium in which RNA is not polyadenylated. *FEMS Microbiology Letters*, *283*(1), 97–103. https://doi.org/10.1111/j.1574-6968.2008.01157.x

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, *239*(4839), 487–491. https://doi.org/10.1126/science.239.4839.487

Sarkar, D., & Fisher, P. B. (2006). Human polynucleotide phosphorylase (hpnpaseold-35): An RNA degradation enzyme with pleiotrophic biological effects. *Cell Cycle*, *5*(10), 1080–1084. https://doi.org/10.4161/cc.5.10.2741

Schneemann, A., Schneider, P. A., Kim, S., & Lipkin, W. I. (1994). Identification of signal sequences that control transcription of Borna Disease Virus, a nonsegmented, negative-strand RNA virus. *Journal of Virology*, *68*(10), 6514–6522. https://doi.org/10.1128/jvi.68.10.6514-6522.1994

Schneider, H., Kaelin, K., & Billeter, M. A. (1997). Recombinant measles viruses defective for RNA editing and V protein synthesis are viable in cultured cells. *Virology*, *227*(2), 314–322. https://doi.org/10.1006/viro.1996.8339

Slomovic, S., Portnoy, V., Liveanu, V., & Schuster, G. (2006). RNA polyadenylation in prokaryotes and organelles; different tails tell different tales. *Critical Reviews in Plant Sciences*, *25*(1), 65–77. https://doi.org/10.1080/07352680500391337

Steinhauer, D. A., Domingo, E., & Holland, J. J. (1992). Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene*, *122*(2), 281–288. https://doi.org/10.1016/0378-1119(92)90216-c

Sutherland, K. A., Collins, P. L., & Peeples, M. E. (2001). Synergistic effects of gene-end signal mutations and the M2-1 protein on transcription termination by respiratory syncytial virus. *Virology*, *288*(2), 295–307. https://doi.org/10.1006/viro.2001.1105

Suttle, C. A. (2005). Viruses in the sea. *Nature*, *437*(7057), 356–361. https://doi.org/10.1038/nature04160

SIB Swiss Institute of Bioinformatics | Disclaimer. *SIB Swiss Institute of Bioinformatics*. SIB Swiss Institute of Bioinformatics | Disclaimer. Retrieved 2022, from https://viralzone.expasy.org/1916

Symmons, M. F., Jones, G. H., & Luisi, B. F. (2000). A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation. *Structure*, *8*(11), 1215–1226. https://doi.org/10.1016/s0969-2126(00)00521-9

Raynal, L. C., & Carpousis, A. J. (1999). Poly(A) polymerase I of escherichia coli: Characterization of the catalytic domain, an RNA binding site and regions for the interaction with proteins involved in mrna degradation. *Molecular Microbiology*, *32*(4), 765–775. https://doi.org/10.1046/j.1365-2958.1999.01394.x

Reuven, N. B., Zhou, Z., & Deutscher, M. P. (1997). Functional overlap of trna nucleotidyltransferase, poly(a) polymerase I, and polynucleotide phosphorylase. *Journal of Biological Chemistry*, *272*(52), 33255–33259. https://doi.org/10.1074/jbc.272.52.33255

Régnier, P., Grunberg-Manago, M., & Portier, C. (1987). Nucleotide sequence of the PNP gene of escherichia coli encoding polynucleotide phosphorylase. homology of the primary structure of the protein with the RNA-binding domain of ribosomal protein S1. *Journal of Biological Chemistry*, *262*(1), 63–68. https://doi.org/10.1016/s0021-9258(19)75888-8

Richter, J. D. (1999). Cytoplasmic polyadenylation in development and beyond. *Microbiology and Molecular Biology Reviews*, *63*(2), 446–456. https://doi.org/10.1128/mmbr.63.2.446-456.1999

Rodamilans, B., Valli, A., Mingot, A., San León, D., Baulcombe, D., López-Moya, J. J., & García, J. A. (2015). RNA polymerase slippage as a mechanism for the production of frameshift gene products in plant viruses of the Potyviridae family. *Journal of Virology*, *89*(13), 6965–6967. https://doi.org/10.1128/jvi.00337-15

Rorbach, J., Bobrowicz, A., Pearce, S., & Minczuk, M. (2014). Polyadenylation in bacteria and organelles. *Polyadenylation*, 211–227. https://doi.org/10.1007/978-1-62703-971-0_18

Rott, R., Zipor, G., Portnoy, V., Liveanu, V., & Schuster, G. (2003). RNA polyadenylation and degradation in cyanobacteria are similar to the chloroplast but different from escherichia coli. *Journal of Biological Chemistry*, *278*(18), 15771–15777. https://doi.org/10.1074/jbc.m211571200

te Velthuis, A. J., Grimes, J. M., & Fodor, E. (2021). Structural insights into RNA polymerases of negative-sense RNA viruses. *Nature Reviews Microbiology*, *19*(5), 303–318. https://doi.org/10.1038/s41579-020-00501-8

Thomas, S. M., Lamb, R. A., & Paterson, R. G. (1988). Two mrnas that differ by two nontemplated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5. *Cell*, *54*(6), 891–902. https://doi.org/10.1016/s0092-8674(88)91285-8

Tomita, K., & Yamashita, S. (2014). Molecular mechanisms of template-independent RNA polymerization by trna nucleotidyltransferases. *Frontiers in Genetics*, *5*. https://doi.org/10.3389/fgene.2014.00036

Valentine, R. C., Thang, M. N., & Grunberg-Manago, M. (1969). Electron microscopy of escherichia coli polynucleotide phosphorylase molecules and polyribonucleotide formation. *Journal of Molecular Biology*, *39*(2), 389–391. https://doi.org/10.1016/0022-2836(69)90326-x

Vidal, S., Curran, J., & Kolakofsky, D. (1990). A stuttering model for paramyxovirus p mrna editing. *The EMBO Journal*, *9*(6), 2017–2022. https://doi.org/10.1002/j.1460-2075.1990.tb08330.x

Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E., & Andino, R. (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature*, *439*(7074), 344–348. https://doi.org/10.1038/nature04388

Villalba, A., Coll, O., & Gebauer, F. (2011). Cytoplasmic polyadenylation and Translational Control. *Current Opinion in Genetics & Development*, *21*(4), 452–457. https://doi.org/10.1016/j.gde.2011.04.006

Whelan, S. P., Barr, J. N., & Wertz, G. W. (2004). Transcription and replication of nonsegmented negative-strand RNA viruses. *Current Topics in Microbiology and Immunology*, 61–119. https://doi.org/10.1007/978-3-662-06099-5_3

Yehudai-Resheff, S., Hirsh, M., & Schuster, G. (2001). Polynucleotide phosphorylase functions as both an exonuclease and a poly(a) polymerase in spinach chloroplasts. *Molecular and Cellular Biology*, *21*(16), 5408–5416. https://doi.org/10.1128/mcb.21.16.5408-5416.2001

Yue, D., Maizels, N., & Weiner, A.M. (1996). CCA-adding enzymes and poly(A) polymerases are all members of the same nucleotidyltransferase superfamily: characterization of the CCA-adding enzyme from the archaeal hyperthermophile Sulfolobus shibatae. *RNA (New York, N.Y.) 2*(9), 895–908.

Zheng, H., Lee, H. A., Palese, P., & García-Sastre Adolfo. (1999). Influenza A virus RNA polymerase has the ability to stutter at the polyadenylation site of a viral RNA template during RNA replication. *Journal of Virology*, *73*(6), 5240–5243. https://doi.org/10.1128/jvi.73.6.5240-5243.1999

Zhuang, Y., Zhang, H., & Lin, S. (2013). Polyadenylation of 18S rrna in algae1. *Journal of Phycology*, *49*(3), 570–579. https://doi.org/10.1111/jpy.12068