*International Journal of Biochemistry Research & Review*

***30(7): 15-59, 2021; Article no.IJBCRR.75573***
***ISSN: 2231-086X, NLM ID: 101654445***

# An Insight into the Proofreading Functions of Multisubunit DNA-Dependent RNA Polymerases and Their Catalytic Mechanism

**Peramachi Palanivelu[1*#]**

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

***Author's contribution***

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

***Article Information***

DOI: 10.9734/IJBCRR/2021/v30i730278
Editor(s):
(1) Dr. Chunying Li, Georgia State University, USA.
Reviewers:
(1) Ioana Stanciu, University of Bucharest, Romania.
(2) Kruti Dave, Ganpat University, India.
(3) F. Solano, University of Murcia, Spain.
Complete Peer review History: http://www.sdiarticle4.com/review-history/75573

***Original Research Article***

***Received 25 August 2021***
***Accepted 03 November 2021***
***Published 09 November 2021***

## ABSTRACT

**Aim:** To analyze the active sites of the proofreading (PR) functions in the multisubunit DNA-dependent RNA polymerases (MSU RNAPs) from prokaryotes, chloroplasts and eukaryotes, and propose a plausible unified catalytic mechanism for these enzymes.
**Study Design:** Data collected on these enzymes from bioinformatics, biochemical, site-directed mutagenesis (SDM), X-ray crystallography and cryo-electron microscopy (cryo-EM) were used for the analyses.
**Methodology:** The protein sequence data of MSU RNAPs from prokaryotes, prokaryotic-types (plant chloroplasts) and eukaryotes 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 from biochemical and SDM experiments, and X-ray crystallographic and cryo-EM data on these enzymes are also used to confirm the possible amino acids involved in the active site of the PR function in these MSU RNAPs
**Results:** All the seven types of MSU RNAPs (I-VII) reported from prokaryotes to eukaryotes were analyzed by the multiple sequence alignment (MSA) software, Clustal Omega, to find out

*[#]Senior Professor & Head (Retd.);*
**Corresponding author: E-mail: ppmkupp@gmail.com;*

conservations among them. The MSA analysis showed many conserved amino acid motifs including small and large peptide regions from the MSU RNAPs of prokaryotes, eukaryotes and plant chloroplasts. Interestingly, the catalytic amino acid and template-binding pairs are highly conserved in all these polymerases, with a few exceptions. Most of them use a basic amino acid (R/K/H) for initiating catalysis and an -YG/FG- pair for template-binding. Some odd type of catalytic amino acids and template-binding pairs are observed in human pathogens, parasites and organisms which cannot ferment sugars. In all the MSU RNAPs, the proposed polymerase catalytic region also possessed three invariant Cs and an invariant H within it. The invariant Cs is shown to bind a zinc atom and proposed to involve in the PR function by excising any misincorporated nucleotide during the transcription process. In the plant-specific MSU RNAPs IV and V, which involve in transcriptional gene silencing in plants, the catalytic and template-binding pairs do not follow the regular distance conservations as observed with other five of the MSU RNAPs. Their polymerase/PR active site regions are similar to RNAP III rather than to RNAP II, as all three make only low molecular weight RNAs.

**Conclusions:** All the known MSU RNAPs possess three invariant Cs and an invariant H embedded within the polymerase active site itself. The three invariant Cs are shown to bind a zinc atom and the invariant H could act as the proton acceptor from a metal-bound water molecule, for initiating excision of the mismatches by a Zn-mediated hydrolysis. Thus, the PR function in MSU RNAPs is integrated within the polymerase active site itself, which is in sharp contrast to the PR functions reported in DNA-dependent DNA polymerases and RNA-dependent RNA polymerases. Therefore, all the seven MSU RNAPs from prokaryotes and eukaryotes are proposed to follow a unified mechanism to excise the mismatches during transcription. The discovery of intrinsic self-correcting RNA transcription mechanism fulfils the missing link in molecular evolution.

*Keywords: Multisubunit RNA polymerases; E. coli; chloroplasts; S. cerevisiae; plant-specific RNA polymerases; Arabidopsis thaliana; human pathogens; zinc-binding site; proofreading active site; catalytic mechanism.*

## 1. INTRODUCTION

MSU RNAPs (EC 2.7.7.6) are key enzymes involved in gene expression in both prokaryotes and eukaryotes and play a crucial role in the flow of genetic information from DNA to proteins through the intermediate transcription process. Therefore, they are one of the essential enzymes found in all living cells and play an important role in copying DNA sequences into RNA molecules. The transcription process forms the first step and a key control point in gene expression and regulation in all organisms. The number and subunit compositions of these MSU RNAPs are varied in different organisms. For example, both eubacteria and archaebacteria contain a single type of MSU RNAP, whereas eukaryotes contain at least five distinct types of MSU RNAPs, viz. RNAP I to RNAP V [1]. Whereas the eubacterial enzymes are composed of 5 different subunits, the eukaryotic enzymes are made up of 12-17 different subunits. Despite such major differences in their compositions, there are striking similarities among the transcriptional mechanisms by these MSU RNAPs across the three major domains of life [1 and references therein]. For example, the amino acid sequences, overall 3D structures and functions of these MSU RNAPs are more or less universally conserved in all organisms like eubacteria, archaebacteria, plants, insects and animals, with small, but significant differences in their active sites, catalytic amino acids, template-binding pairs and zinc-binding sites. Errors in eukaryotic transcription process can potentially lead to aberrant gene products, which could eventually lead to various diseases in humans, including cancer. For example, one major class of transcription error, known as transcriptional slippages, are implicated in the development of a wide variety of diseases, like colon cancer, non-familial Alzheimer's, Down's syndrome, etc. [2]. Recently, the *in vitro* transcribed mRNAs have come into focus as a potential new class of drugs known as 'mRNA therapeutics' to deliver genetic information through mRNAs to correct the malfunction(s) [3,4] and also potential vaccines against cancer and viral infections, e.g., SARS-CoV-2 [5,6]. In fact, many anti-cancer drugs act by inhibiting the transcription step itself. Therefore, understanding the catalytic mechanism and regulation of these MSU RNAPs will be useful in designing novel mRNA-based drugs for many diseases.

### 1.1 Proofreading Functions in MSU RNAPs

As transcription is the first and crucial step in gene expression, the mechanism of transcription is highly conserved in all organisms. It is accomplished mainly by the MSU RNAPs both in prokaryotes and eukaryotes. As the RNAPs are the key enzymes in creating an equivalent copy of RNA from the DNA sequence, they maintain a very high fidelity during the transcription process. Generally, the MSU RNAPs misincorporates only one wrong nucleotide per ~100,000 bases. Such mismatched nucleotides during transcription usually lead to conformational changes at the polymerase active site resulting in stalling of the RNAP. Fidelity of DNA replications by DNA polymerases (DNAPs) is regulated by a well established PR mechanism. However, existence of a similar mechanism for RNA synthesis is not well understood [7,8]. The polymerization and PR active sites of replicative DNA polymerases are located either on different domains on the same polypeptide or in a different subunit associated with the replicase [1,9]. However, such distinct PR domain(s) or subunit(s) are not demonstrated in MSU RNAPs. Therefore, it has been proposed that in MSU RNAPs, the transcriptional fidelity could be achieved by a different mechanism, (i.e.), it could be achieved in two steps. In the first step, it could discriminate deoxynucleoside triphosphates (dNTPs) from nucleoside triphosphates (NTPs) and allow only the NTPs at the polymerization site. In the second step, they could also employ a PR function to repair any mismatch. The second step in these MSU RNAPs has been proposed to be possibly achieved by a PR site integrated within the polymerase active site itself. Thus, any mismatch could be repaired by an intrinsic PR mechanism and hence, there may not be any separate domains or subunits to perform the function as in the replicative polymerases [10,11]. Interestingly, huge volume of genomic and protein sequence data that are available for most of the prokaryotic and eukaryotic MSU RNAPs could be effectively utilized to analyze the structure-function relationships and catalytic mechanism of these key enzymes. Therefore, in this communication, the PR function in these MSU RNAPs is analyzed in detail and reported.

## 2. MATERIALS AND METHODS

The protein sequence data of MSU RNAPs from prokaryotes, eukaryotes and plant chloroplasts were obtained from PUBMED and SWISS-PROT databases. The advanced version of Clustal Omega was used for protein sequence analysis. The data already available from bioinformatics methods, biochemical and SDM experiments and X-ray crystallographic and cryo-EM analysis on these enzymes were also used to confirm the possible amino acids involved in the PR exonuclease active site of these MSU RNAPs.

## 3. RESULTS AND DISCUSSION

### 3.1 Analysis of PR Active Site in the MSU RNAPs from Prokaryotes

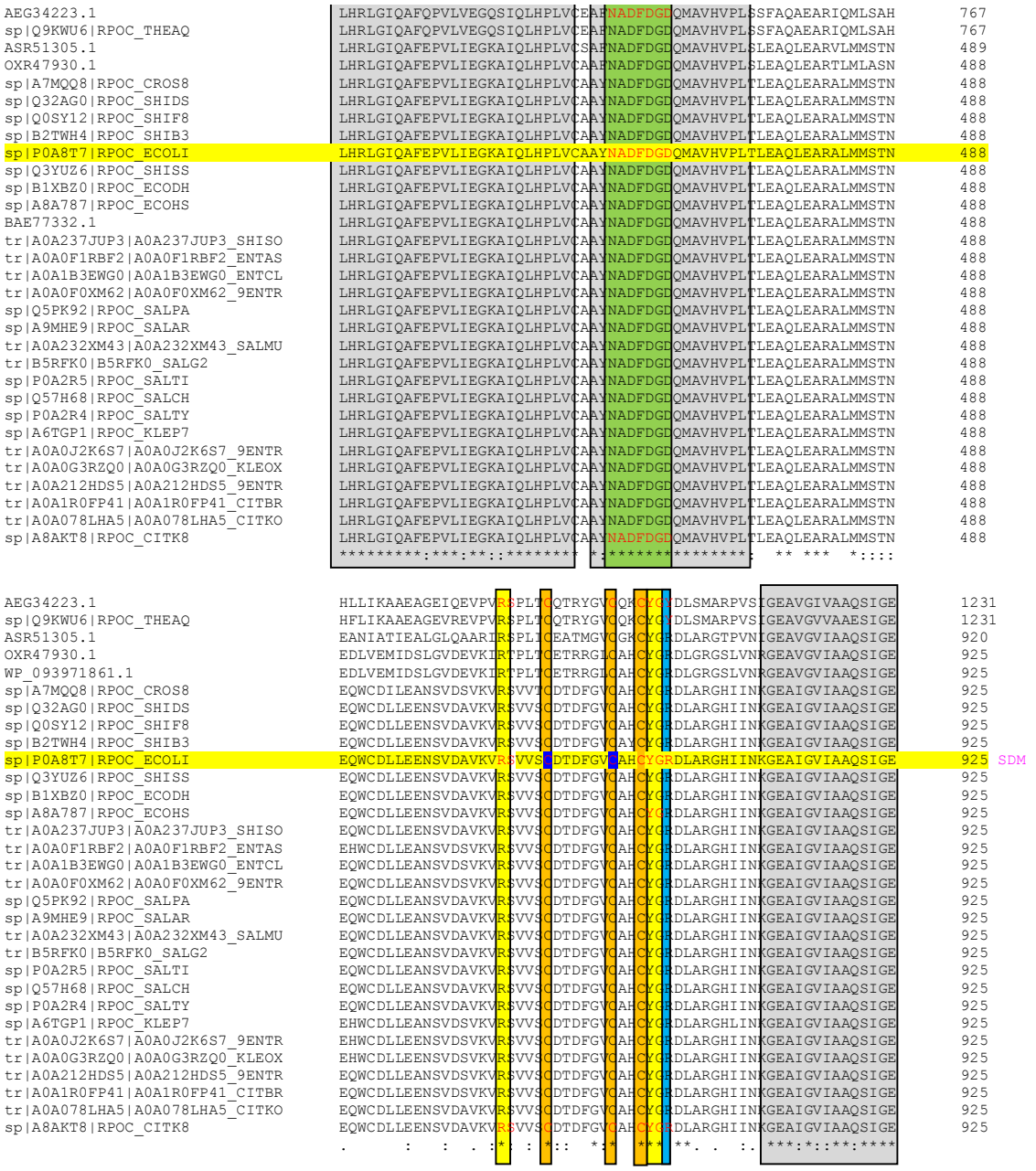
The MSA of the elongation subunits β' of the MSU RNAPs from prokaryotes is shown in Fig. 1. The most well studied among them is the RNAP of *E. coli*. The *E coli* RNAP is a MSU enzyme with a molecular mass of ~410 kDA. The core enzyme consists of four different subunits and organized into $\alpha_2\beta\beta'\omega$ in bacteria. In addition to the core enzyme, the RNAP also requires an additional subunit, viz. the sigma factor (σ) which essentially involves in promoter recognitions [8]. In bacteria, the β subunit involves in the initiation of RNA synthesis and the β' subunit is responsible for further elongation and completion of the RNA transcription cycle. Thus, the β and β' subunits are complementary to each other in the RNA synthesis. In fact, the β subunit involves in synthesis of short primers for initiation of RNA synthesis, and the β' is responsible for further extension of the primers and production of the complete RNA chains, ready for translation. Therefore, it is important for the elongation β' subunit to harbour both the polymerization and the PR active sites for an error-free transcription process. The β' elongation subunits from various bacterial sources were analyzed by MSA and the results are shown in Fig. 1. (The *E. coli* β' subunit, the template-binding and catalytic pairs are highlighted in yellow; the $Mg^{2+}$ binding site is highlighted in light green and the possible $Zn^{2+}$ binding motif $–Cx_6CxxC-$ is highlighted in orange).

#### 3.1.1 Identifications of $Mg^{2+}$- and $Zn^{2+}$-Binding Sites in the MSU RNAPs from prokaryotes

The MSA shows a highly conserved Cs with the $Zn^{2+}$-binding pattern $–Cx_6Cx_2C-$ within the polymerase active site region (Fig. 1). The presence of a $Zn^{2+}$ binding site is further confirmed by X-ray crystallographic analysis of the bacterial (*E. coli* and *T. thermophilus)* MSU RNAPs [9] and references therein. Zhang et al

[12] found that the MSU RNAP from *T. aquaticus* showed an unusual $Zn^{2+}$-binding motif in the β' elongation subunit of the enzyme. They reported that the three C residues are arranged in a sequence ($^{1194}CX_6CX_2C$), similar to a typical Zn-binding motif. The fourth C participating in the $Zn^{2+}$ chelation is $C^{1113}$, (i.e.,) 82 residues away. An absolute conservation of the four C residues in all β' subunits from prokaryotes further support this finding ($C^{888}$, $C^{895}$ and $C^{898}$ - numbering from *E. coli* β'). Moreover, the complete conservation of the zinc binding motif in the active sites of the elongation subunits in the bacterial MSU RNAPs by MSA analysis further confirms this finding (Fig.1). This finding was further corroborated by

CLUSTAL O (1.2.4) MSA of the elongation subunits, β' of bacterial MSU RNAPs
(Only the $Mg^{2+}$-binding and polymerization regions are shown here).



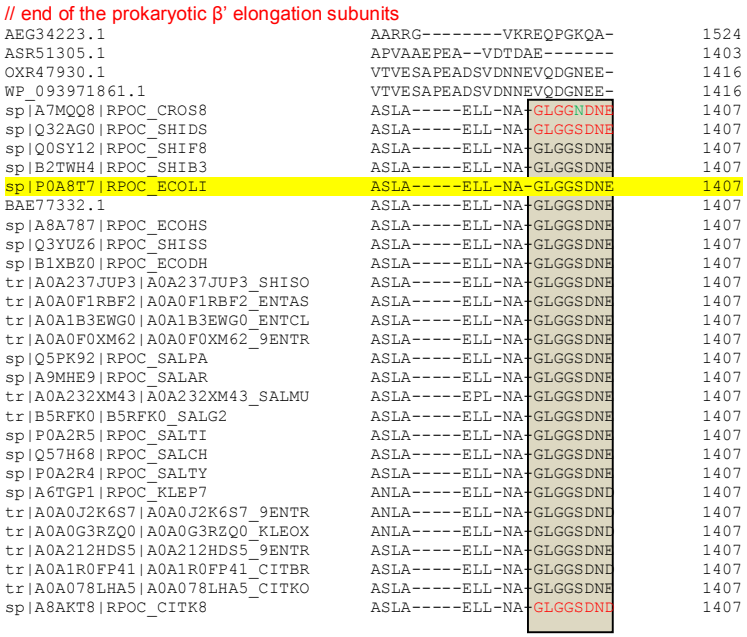



**Fig. 1. MSA of the β' elongation subunits of the MSU RNAPs from different bacteria**

*AEG34223.1 Thermus thermophilus*
*ASR51305.1 Blastomonas fulva*
*WP_093971861.1 Pusillimonas* sp. T2
*Q32AG0|RPOC_SHIDS Shigella dysenteriae,*
*B2TWH4|RPOC_SHIB3 Shigella boydii,*
*A8A787|RPOC_ECOHS Escherichia coli* O9:H4
*B1XBZ0|RPOC_ECODH Escherichia coli* (K12)
*A0A0F1RBF2_ENTAS Enterobacter asburiae*
*A0A0F0XM62_9ENTR Enterobacter kobei*
*A9MHE9|RPOC_SALAR Salmonella arizonae*
*B5RFK0_SALG2 Salmonella gallinarum*
*Q57H68|RPOC_SALCH Salmonella choleraesuis*
*A6TGP1|RPOC_KLEP7 Klebsiella pneumonia*
*A0A0G3RZQ0_KLEOX Klebsiella oxytoca*
*A0A1R0FP41_CITBR Citrobacter braakii*
*A8AKT8|RPOC_CITK8 Citrobacter koseri* (K8)
*Q9KWU6|RPOC_THEAQ Thermus aquaticus*
*OXR47930.1 Pusillimonas* sp.
*A7MQQ8|RPOC_CROS8 Cronobacter sakazakii*
*Q0SY12|RPOC_SHIF8 Shigella flexneri serotype 5b*
*P0A8T7|RPOC_ECOLI Escherichia coli* (strain K12)
*Q3YUZ6|RPOC_SHISS Shigella sonnei*
*A0A237JUP3_SHISO Shigella sonnei*
*A0A1B3EWG0_ENTCL Enterobacter cloacae*
*Q5PK92|RPOC_SALPA Salmonella paratyphi A*
*A0A232XM43_SALMU Salmonella muenchen*
*P0A2R5|RPOC_SALTI Salmonella typhi*
*P0A2R4|RPOC_SALTY Salmonella typhimurium* (LT2)
*A0A0J2K6S7_9ENTR Klebsiella michiganensis*
*A0A212HDS5_9ENTR Citrobacter* sp. 86
*A0A078LHA5_CITKO Citrobacter koseri*

an SDM experiment where the first two Cs modification (marked in dark blue) lead to the loss of the enzyme activity [13]. It is interesting to note that similar intrinsic zinc binding PR site is not observed in the initiation subunits β, as they make only small primers for RNA elongation (-$^{539}$TRER$^{-6}$AGFEVRDVHPTHYGR$^{557}$- numbering from *E. coli* β subunit) [8].

In addition to the $Zn^{2+}$-binding motif, X-ray crystallographic analysis of the *T. aquaticus* RNAP also showed a $Mg^{2+}$-binding site. The $Mg^{2+}$ was chelated at an absolutely conserved –NADFDGD- motif in the β' subunits and that apply to all other prokaryotic RNAPs (e.g., **–**$^{458}$**N**ADFDGD**-** (numbering from *E. coli* β' subunit) [14]. The $Mg^{2+}$-binding site (site A) was further confirmed by SDM experiments. Substitution of the invariant Ds by A (D→A) gave rise to a dominant lethal phenotype and showed no detectable enzyme activity [13]. Sydow and Cramer have suggested that a mismatch during transcription could cause loss of the catalytic metal ion A ($Mg^{2+}$) which might give way for the metal ion B ($Zn^{2+}$) to perform the PR activity [11]. These results strengthen the bifunctional, "tunable" mode of active site in MSU RNAPs [10].

### 3.1.2 Proposed mechanism for the PR and polymerization reactions in bacterial MSU RNAPs

As suggested elsewhere, the PR function is found to be integrated into the polymerase active site itself in MSU RNAPs. Thus, the polymerization is achieved using the metal ion $Mg^{2+}$ and the PR function is achieved by the second metal ion $Zn^{2+}$. Modeling of the substrate NTP bound to the *T. thermophilus* RNAP active site suggests that $N^{458}$ (numbering from *E. coli* RNA polymerase) within a highly conserved sequence motif **-**$^{458}$NA**D**F**D**G**D**$^{464}$**-** that includes the catalytic Asp triad ($D^{460}$, $D^{462}$ and $D^{464}$) which could mediate specific recognition of the O2 ribose atom and NTP selection [15,12] Both the catalytic metal ions are coordinated by a set of completely conserved amino acids in all the MSU RNAPs. Fig. 2A shows the proposed active amino acids of the PR exonuclease in the MSU RNAP from *E. coli*.

During a mismatch the polymerase pauses, giving way for the PR exonuclease active site, embedded within the polymerases active site itself, to excise the wrongly inserted nucleotide. The RNAP PR exonuclease unlike the PR exonucleases of DNAPs and RNA-dependent RNAPs, backtracks one nucleotide from the mismatch and make a cut on the penultimate base removing a dinucleotide [16] (Fig. 2B).

## 3.2 Analysis of PR Active Site in the MSU RNAPs from Plant Chloroplasts

MSU RNAP from plant chloroplasts is also a well characterized one. It belongs to prokaryotic-type and is very similar to eubacterial MSU RNAP discussed elsewhere. The MSU RNAP from plant chloroplasts is encoded by the plastids and hence, also known as plastid-encoded polymerase (PEP). The MSU enzyme from chloroplasts is structurally similar to their eubacterial counterparts, except having an additional subunit, **β"** and hence the chloroplast enzyme's subunit composition is α, β, β′, **β"** and ω. [9]. (Chloroplasts also possess another RNAP which is encoded by the nucleus (NEP) and imported into the chloroplasts. It is structurally unrelated to PEP and belongs to the single subunit (SSU) RNAP types and is very similar to the SSU RNAPs of bacteriophages T3, T7, SP6, etc. in structure and function). In chloroplast MSU RNAPs, the β is involved in the initiation of transcription like in prokaryotes and both the β′ and β" subunits involve in the elongation process. Whereas the $Mg^{2+}$-binding site is found in the β′ subunit, the polymerization active site is found in the β"' subunit. MSAs show the $Mg^{2+}$-binding site from β′ subunits and the polymerization active site from β" subunits of the RNAPs of chloroplasts from various plant species (Figs. 3a, 3b).

Figure 3a shows the $Mg^{2+}$-binding site in the β' subunits from various plant sources. It shows a completely conserved $Mg^{2+}$-binding site (**-**NA**D**F**D**G**D-**) with three invariant Ds as in the prokaryotic types, suggesting its possible prokaryotic origin. The maize enzyme is highlighted in yellow and the $Mg^{2+}$-binding site is in light green.

Figure 3b shows the polymerization active site region from the β'' subunits of the MSU RNAPs of chloroplasts from various plant sources. The template-binding and catalytic pairs are highlighted in yellow and the $Zn^{2+}$-binding Cys residues are highlighted in orange. It shows very similar polymerization active site with the zinc-binding motif (-$CX_6CX_2C$-) like the prokaryotic type with similar template-binding and catalytic pairs, suggesting its possible origin.

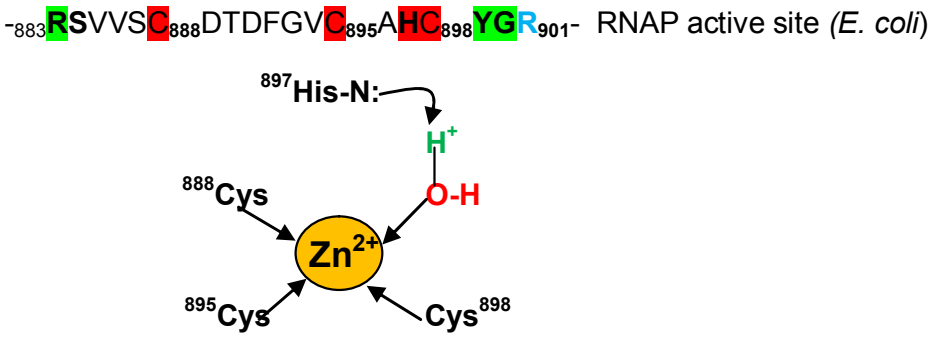



**Fig. 2A. Proposed active site in prokaryotic MSU RNAPs (numbering from *E. coli* β' subunit)**

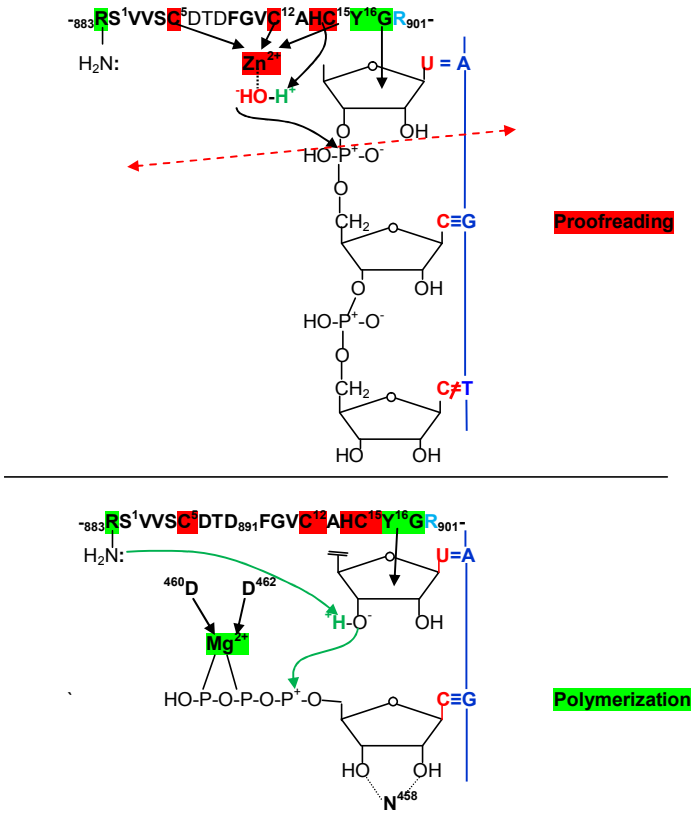



**Fig. 2B. A simplified, schematic diagramme showing the proposed reactions of the PR and polymerization active sites in the mismatch repair and polymerization processes in bacterial MSU RNAPs (numbering from *E. coli* β' subunit).**

CLUSTAL O (1.2.4) MSA of the elongation subunits β' of MSU RNAPs from plant chloroplasts (Only the $Mg^{2+}$-binding region of the β' subunits is shown here)

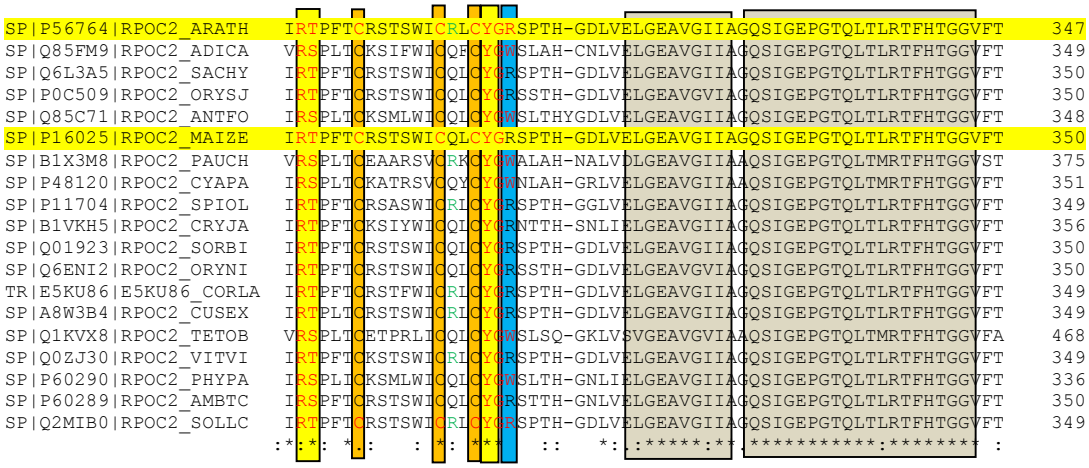
```
SP|P0C506|RPOC1_ORYSJ RAPTLHRLGIQAFQPTLVEGRTICLHPLVCKGFNADFDGDQMAVHLPLSLEAQAEARLLM 513
SP|A7M957|RPOC1_CUSRE RAPTLHRLGIQAFQPVLVEGHVLCLHPLVCKGFNADFDGDQMAVHVPLSLEAQAEARLLM 520
SP|B1X3M9|RPOC1_PAUCH RAPTLHRLGIQAFEPKLVDGRAICLHPLVCPAFNADFDGDQMAVHVPLSIEAQTEARILM 504
SP|Q1XDN6|RPOC1_PYRYE RAPTLHRLGIQAFEPILVEGRAIKLHPLVCPAFNADFDGDQMAVHVPLSLEAQAEARLLM 496
SP|P42080|RPOC1_CYAPA RAPTLHRLSIQAFEPILVEGRAICLHPLVCPAFNADFDGDQMAVHVPLSLEAQTEARLLM 495
SP|Q4G3A6|RPOC1_EMIHU RAPTLHRLGIQSFEPILVSGRAIKLHPLVCPAFNADFDGDQMAVHIPLSLEAQSEARLLM 484
SP|O19897|RPOC1_CYACA RAPTLHRLGIQAFDPVLVDGRAIRLHPLVCPAFNADFDGDQMAVHIPLSIEAQTEARLLM 492
SP|P51251|RPOC1_PORPU RAPTLHRLGIQAFEPILVEGRAIKLHPLVCPAFNADFDGDQMAVHVPLSLEAQAEARLLM 496
SP|Q6B8R7|RPOC1_GRATL RAPTLHRLGIQAFEPILVEGRAIKLHPLVCPAFNADFDGDQMAVHVPLSLEAQAEARLLM 496
SP|P56763|RPOC1_ARATH RAPTLHRLGIQSFQPILVEGRTICLHPLVCKGFNADFDGDQMAVHVPLSLEAQAEARLLM 513
SP|P16024|RPOC1_MAIZE RAPTLHRLGIQAFQPTLVEGRTICLHPLVCKGFNADFDGDQMAVHLPL SLEAQAEARLLM 513
SP|Q85FM8|RPOC1_ADICA RAPTLHRLGIQAFQPLLIEGRAIRLHPLVRGGFNADFDGDQMAVHVPLSVEAQIEARLLM 513
SP|Q85CL6|RPOC1_ANTFO RAPTLHRLGIQAFEPILVEGRAIRLHPLVCAGFNADFDGDQMAVHIPLSLEAQVEARLLM 513
SP|P42079|RPOC1_SYNE7 RAPTLHRLGIQAFEPILVEGRAICLHPLVCPAFNADFDGDQMAVHVPLSIEAQAEARMLM 490
SP|P11705|RPOC1_SPIOL RAPTLHRLGIQAFQPILVEGRAICLHPLVCKGFNADFDGDQMAVHVPLSLEAQAEARLLM 513
SP|P14563|RPOC1_NOSCO RAPTLHRLGIQSFEPILVEGRAICLHPLVCPAFNADFDGDQMAVHVPLSLESQAEARLLM 487
SP|Q2MIA9|RPOC1_SOLLC RAPTLHRLGIQAFQPVLVEGRAICLHPLVCKGFNADFDGDQMAVHVPLSLEAQVEARLLM 513
SP|Q6ENI3|RPOC1_ORYNI RAPTLHRLGIQAFQPTLVEGRTICLHPLVCKGFNADFDGDQMAVHILPLSLEAQAEARLLM 513
SP|P46819|RPOC1_SINAL RAPTLHRLGIQSFQPILVEGRTICLHPLVCKGFNADFDGDQMAVHVPLSLEAQAEARLLM 521
SP|P58131|RPOC1_EUGLO RAPTLHRLNIQAFQPKLTIGKSIKLHPLVCSAFNADFDGDQMGVHIPLSLKAQAEARNIL 64
SP|O78484|RPOC1_GUITH RAPTLHRLGIQAFEPILVEGRAIKLHPLVCPAFNADFDGDQMAVHIPLSLEAQAEARMLM 490
SP|Q2VEI5|RPOC1_SOLTU RAPTLHRLGIQAFQPVLVEGRAICLHPLVCKGFHADFDGDQMDFKVPLSLEAQVEARLLM 519
SP|A6MVX3|RPOC1_RHDSA RAPTLHRLGIQAFEPILVEGRAIKLHPLVCPAFNADFDGDQMAVHIPLSLEAQAEARLLM 490
                       ********.**:*:* *  *: : ***** .*:******** .::***::* *** ::
```

```
// End of β' initiation subunits
SP|P0C506|RPOC1_ORYSJ FSHMNLLSPAIGDPICVPTQDMLIGLYVLTIGNRRGICANRYNSCGNYPNQKVNYNNN-N 572
SP|A7M957|RPOC1_CUSRE FSHMNLLSPAIGDPISVPTQDMLIGLYVLTSDNRRDICTNRYTKCNIQTLQTKSSDSSNS 580
SP|B1X3M9|RPOC1_PAUCH LASANILSPATGEPVITPSQDMVLGIYYLTSSKPAQT---------------------- 541
SP|Q1XDN6|RPOC1_PYRYE LAPHNFLSPATGQPIIMPSQDMVLGCYYLTANNPSQQ---------------------- 533
SP|P42080|RPOC1_CYAPA LASNNLLSPATGQPIVTPSQDMVLGCYYLTVDNLKNQ---------------------- 532
SP|Q4G3A6|RPOC1_EMIHU LAPNNFLSPATGDAILTPSQDMVLGCFYLTANNPSQQ---------------------- 521
SP|O19897|RPOC1_CYACA LAPNNFLSPATGQPIITPSQDMVLGCYYLTNNNLANQ---------------------- 529
SP|P51251|RPOC1_PORPU LAPHNFLSPATGQPIIMPSQDMVLGCYYLTANNPSQQ---------------------- 533
SP|Q6B8R7|RPOC1_GRATL LAPHNFLSPATGQPILMPSQDMVLGCYYLTTYNPAAI---------------------- 533
SP|P56763|RPOC1_ARATH FSHMNLLSPAIGDPISVPTQDMLIGLYVLTSGTRRGICANRYNPCNRKNYQNERI----Y 569
SP|P16024|RPOC1_MAIZE FSHMNLLSPAIGDPICVPTQDMLIGLYVLTIGNRLGICANRYNSCGNSPNKKVNYNNN-N 572
SP|Q85FM8|RPOC1_ADICA FSHLNLLSPATGDPVSVPSQDMLLGLYALTIESRQGIYRNRHLGFIDRVN---------- 563
SP|Q85CL6|RPOC1_ANTFO FSHTNLLSPVTGNPVSVPSQDMLLGIYVSTIRSNRGIYQNQYHPYDYRNK---------- 563
SP|P42079|RPOC1_SYNE7 LASGNILSPATGQPIVTPSQDMVLGCYYLTAENPGAQ---------------------- 527
SP|P11705|RPOC1_SPIOL FSHMNLLSPAIGDPISVPTQDMLIGLYILTSGNRRGICANRYNPWNHKTYQNERI----D 569
SP|P14563|RPOC1_NOSCO LASNNILSPATGKPIITPSQDMVLGAYYLTAENPGAT---------------------- 524
SP|Q2MIA9|RPOC1_SOLLC FSHMNLLSPAIGDPISVPTQDMLIGLYVLTSGNHRGICVNRYNPCNRRNYQNQKRS---D 570
SP|Q6ENI3|RPOC1_ORYNI FSHMNLLSPAIGDPICVPTQDMLIGLYVLTIGNRRGICANRYNSCGNYPNQKVNYNNN-N 572
SP|P46819|RPOC1_SINAL FSHMNLLSPAIGDPISVPTQDMLIGLYVLTSGTRRGICANRYNPCNRKNYQNERI----Y 577
SP|P58131|RPOC1_EUGLO ISINNCNSLKNGDPNILPSQDIILGCYFSNIENCNLL---------------------- 501
SP|O78484|RPOC1_GUITH LAPYNFLSPATGDPIIMPSQDMVLGCYYLTAENPSQQ---------------------- 527
SP|Q2VEI5|RPOC1_SOLTU FSHMNLLSPAIGDPISVPTQDMLIGLYVLTSGNHRGICVNRYNPCNRRNYQNQKRS---D 576
SP|A6MVX3|RPOC1_RHDSA LAPYNFLSPATGEPIIMPSQDMVLGCYYLTTHNPSQQ---------------------- 527
                      ::  *  *   *.    *:**:::* :  .  .
```

**Fig. 3a. MSA of β' initiation subunits of MSU RNAPs of chloroplasts from plant sources**

*P0C506|RPOC1_ORYSJ Oryza sativa* subsp. *JaponicaA7M957|RPOC1_CUSRE Cuscuta reflexa*
*B1X3M9|RPOC1_PAUCH Paulinella chromatophora* *Q1XDN6|RPOC1_PYRYE Pyropia yezoensis*
*P42080|RPOC1_CYAPA Cyanophora paradoxa* *Q4G3A6|RPOC1_EMIHU Emiliania huxleyi*
*O19897|RPOC1_CYACA Cyanidium caldarium* *P51251|RPOC1_PORPU Porphyra purpurea*
*Q6B8R7|RPOC1_GRATL Gracilaria tenuistipitata* *P56763|RPOC1_ARATH Arabidopsis thaliana*
*P16024|RPOC1_MAIZE Zea mays* *Q85FM8|RPOC1_ADICAAdiantum capillus-eneris*
*Q85CL6|RPOC1_ANTFO Anthoceros formosae* *P42079|RPOC1_SYNE7Synechococcus elongatus*
*P11705|RPOC1_SPIOL Spinacia oleracea* *P14563|RPOC1_NOSCO Nostoc commune*
*Q2MIA9|RPOC1_SOLLC Solanum lycopersicum* *Q6ENI3|RPOC1_ORYNI Oryza nivara*
*P46819|RPOC1_SINAL Sinapis alba* *P58131|RPOC1_EUGLO Euglena longa*
*O78484|RPOC1_GUITH Guillardia theta* *Q2VEI5|RPOC1_SOLTU Solanum tuberosum*
*A6MVX3|RPOC1_RHDSA Rhodomonas salina*

CLUSTAL O (1.2.4) MSA of the β'' elongation subunits of MSU RNAPs from chloroplasts
(Only the polymerization region of the elongation subunits is shown here)



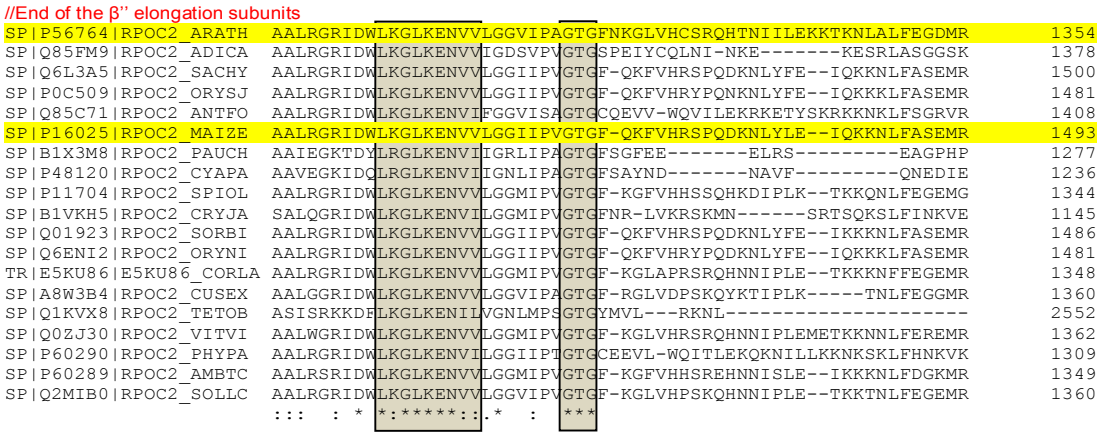



**Fig. 3b. MSA of β'' subunits MSU RNAPs from chloroplasts**

*P56764|RPOC2_ARATH Arabidopsis thaliana*
*Q6L3A5|RPOC2_SACHY Saccharum hybrid*
*Q85C71|RPOC2_ANTFO Anthoceros formosae*
*B1X3M8|RPOC2_PAUCH Paulinella chromatophora*
*P11704|RPOC2_SPIOL Spinacia oleracea*
*Q01923|RPOC2_SORBI Sorghum bicolor*
*E5KU86|E5KU86_CORLA Corynocarpus laevigatus*
*Q1KVX8|RPOC2_TETOB Tetradesmus obliquus*
*P60290|RPOC2_PHYPA Physcomitrella patens*
*Q2MIB0|RPOC2_SOLLC Solanum lycopersicum*
*Q85FM9|RPOC2_ADICA Adiantum capillus-veneris*
*P0C509|RPOC2_ORYSJ Oryza sativa subsp. japonica*
*P16025|RPOC2_MAIZE Zea mays*
*P48120|RPOC2_CYAPA Cyanophora paradoxa*
*B1VKH5|RPOC2_CRYJA Cryptomeria japonica*
*Q6ENI2|RPOC2_ORYNI Oryza nivara*
*A8W3B4|RPOC2_CUSEX Cuscuta exaltata*
*Q0ZJ30|RPOC2_VITVI Vitis vinifera*
*P60289|RPOC2_AMBTC Amborella trichopoda*

## 4. PR FUNCTION IN THE MSU RNAPS FROM EUKARYOTES

Unlike in the prokaryotes where a single RNAP transcribes all the cellular RNAs like mRNAs, rRNAs, tRNAs, in eukaryotes at least 3 distinct RNAPs, viz. RNAP I, RNAP II and RNAP III transcribe the three major RNA species. All the three polymerases are MSU RNAPs and catalyze DNA-dependent RNA synthesis and are localized in the nucleus. RNAP I synthesize the rRNAs, whereas the RNAP II and RNAP III synthesize mRNAs and tRNAs, respectively. In addition to the three major polymerases, two more MSU RNAPs, viz. the RNAP IV and V are reported from plant sources and they mainly involve in synthesizing small interfering RNAs (siRNAs) for gene silencing in plants [17]. Table 1 shows the composition of the major eukaryotic MSU RNAPs, their compositions and functions in the cell. The initiation and elongation subunits are distinct in all three RNAPs. The unique subunits and common subunits for all the three RNAPs are shown in red and green, respectively. The largest subunit, also known as the elongation subunit with a molecular mass of ~160 kDa (which is functionally equivalent to prokaryotic β') is named; RP**A1**; RP**B1**; RP**C1** and the second-largest subunit, also known as the initiation subunit with a molecular mass of ~150 kDa (which is functionally equivalent to prokaryotic β); is named RP**A2**; RP**B2**; RP**C2** from RNAP I, RNAP II and RNAP III, respectively.

### 4.1 PR Function in the Eukaryotic MSU RNAP I

The RNAP I is localized in the nucleolus sub-compartment of the nucleus and involve in the transcription of precursor rRNAs (pre-rRNA). In yeast, RNAP I activity accounts for up to 60% of all nuclear transcription, and the product, the rRNA accounts for up to 80% of the total cellular RNAs [20]. The common cellular rRNAs like 5.8S, 18S, 28S rRNAS are derived from the 45S pre-rRNA by the pre-rRNA processing. These 3 rRNAs together with the 5S rRNA synthesized by the RNAP III, comprise the enzymatic and structural components of the ribosomes. The transcription of the rRNA genes leads to the synthesis of the several millions copies of ribosomes needed in actively growing cells. Therefore, their synthesis is the first step in ribosome biogenesis and regulation of cell growth. In fact, in the rapidly growing cancer cells, high levels of rRNA synthesis are always maintained. Thus, the up- and down-regulations of rRNA transcription play an important role in oncogenesis.

**Table 1. Subunit compositions and functions of the three major eukaryotic polymerases**

| Subunit structure | RNAP I | RNAP II | RNAP III* |
|---|---|---|---|
| Unique core subunits ($\alpha_2\beta\beta$'-like) | A190<br>A135<br>AC40<br>AC19<br>A12.2 | Rpb1 (β')**<br>Rpb2 (β)<br>Rpb3 (α)<br>Rpb11<br>Rpb9 | C160<br>C128<br>AC40<br>AC19<br>C11 |
| Common subunits | Rpb5<br>Rpb6<br>Rpb8<br>Rpb10<br>Rpb12 | Rpb5<br>Rpb6 (ω)<br>Rpb8<br>Rpb10<br>Rpb12 | Rpb5<br>Rpb6<br>Rpb8<br>Rpb10<br>Rpb12 |
| Stalk sub-complex (A14+A43) | A14<br>A43 | Rpb4<br>Rpb7 | C17<br>C25 |
| TFIIF- like sub-complex (A49+A34.5) | A49<br>A34.5 | TFIIF-α<br>TFIIF-β | C37<br>C53 |
| **Total No. of subunits** | **14** | **12** | **17** |
| **Products** (45S RNA→28S, 5.8S & 18S) , | pre-rRNAS | pre-mRNAs, 5 snRNAs^, SnoRNAs & microRNAs | pre-tRNAs, 5S, 7S RNAs$ & U6-snRNA |
| Sensitivity to α-Amanitin | Nil | High (1 µg/ml) | Moderate (10 µg/ml) |
| Sensitivity to Actinomycin-D# | 0.05 µg/ml | 0.5 µg/ml | 5.0 µg/ml |

*Adapted from [18,19]*
*Current subunit nomenclatures of the RNAPs I, II and III are, RPA1-A14; RPB1-B12; RPC1-C17*
**Pol III also possesses a trimeric sub-complex made up of C82-C34-C31.*
***The largest subunit of RNAP II (Rpb1) possesses a unique CarboxyTerminal Domain (CTD)*
*^U1-U5 of ~200 bases, involves in the formation of spliceosomes*
*$7S RNA from the signal recognition particle (SRP), which is involved in the transport of proteins into the endoplasmic reticulum.*

The RNAP I has a molecular mass of ~600 kDa and made up of 14 subunits [21]. The subunits Rpb5, Rpb6, Rpb8, Rpb10, and Rpb12 are identical in all three polymerases. The two large RNAP I subunits, viz. A190 and A135 are similar in function to the RNAP II subunits Rpb1 (= prokaryotic β') and Rpb2 (= prokaryotic β), respectively [20] and references therein] (Table 1).

The RNAP I from yeast is extensively studied by both cryo-EM and X-ray crystallography [18]. The cryo-EM structure of the complete 14-subunit core enzyme from yeast at 12A° exhibited that the RNAP I showed a strong intrinsic 3'-RNA cleavage activity as compared to RNAP II and RNAP III, which apparently enables rRNA PR activity and end trimming. Furthermore, incubation of the backtracked complex with $Mg^{2+}$ ions led to efficient shortening of the RNA from the 3'-end. An RNAP I variant, lacking residues79–125 of A12.2 subunit = Rpb9 was totally inactive in RNA cleavage, but bound to the nucleic-acid scaffold and retained elongation activity, suggesting the C-terminal domain in the involvement of other subunits in RNA I PR function. Interestingly, the conserved polymerase active site of RNAP I was capable of RNA cleavage in the absence of cleavage stimulatory factors. Thus, the intrinsic RNA cleavage activity apparently enables rRNA 3' trimming and PR activities to prevent any possible errors in rRNAs [18]. The crystal structure of RNAP I from the yeast, *S. cerevisiae*, at 2.8A° resolution with all its 14 subunits was studied by Engel et al [20]. The yeast RNAP I structure reveals the 10-subunit RNAP I core and the sub-complexes A49–A34.5 and A14–A43 on opposite sides. Thus, a composite active site of RNAP I and RNAP III enables efficient PR and termination.

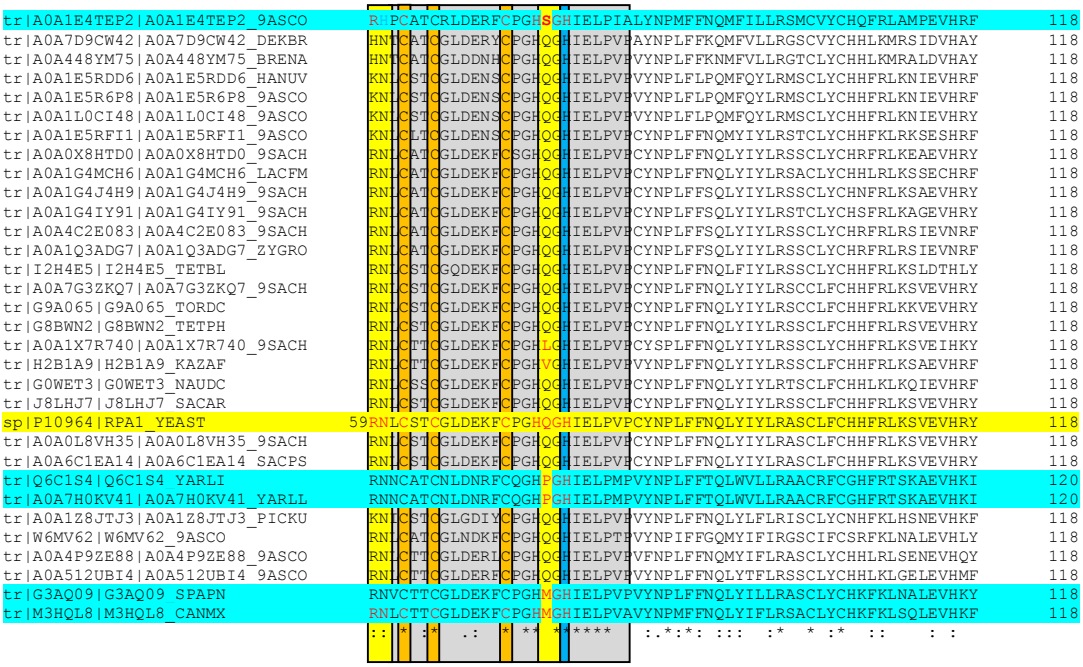
Figure 4 shows the MSA analysis of the elongation subunits of RNAP I from different yeasts. The *S. cerevisiae* sequence, template-binding and the catalytic pairs are highlighted in yellow and the 3 invariant Cs (proposed $Zn^{2+}$-binding site) are highlighted in orange. Though the catalytic pair is almost completely conserved

in all, the template-binding pairs are markedly different as compared to other MSU RNAPs discussed above. Most of them use –QG- (with a few exceptions) instead of the regular -YG/FG- pair. Human pathogens and yeasts, which could not ferment sugars showed altogether different sets of template-binding pairs are discussed in Table 2. However, the $Mg^{2+}$-binding site with three Ds is completely conserved in all (highlighted in light green). The human pathogens and yeasts, which could not ferment sugars are highlighted in light blue. All the template-binding pairs are followed by an invariant H (Fig. 4).

Figure 5 shows the MSA analysis of the elongation subunits of RNAP I from higher fungi. The catalytic pair is completely conserved in all but with a basic amino acid H, instead of the usual R/K, which is followed by an invariant hydrophobic amino acid (V/I). The higher fungal group also uses an uncommon template-binding pairs, e.g., most of them use a -TG- pair instead of the regular –YG/FG- pair and some of them use -PG- or -NG- pairs showing a great diversity in the template-binding pairs among the higher fungi. It is interesting to note that the higher fungi that use the unusual template-binding pairs like –NG-/-PG- are having something in common, (i.e.), most of them are pathogenic, e.g., *Talaromyces marneffei, Ajellomyces capsulatus, and Blastomyces dermatitidis are* human pathogens [22] The *Talaromyces cellulolyticus* is a celluloytic fungus (Table 2). However, the $Mg^{2+}$-binding motif is completely conserved in all of them (Fig. 5).

Figure 6 shows the MSA of the elongation subunits of RNAP I from various plant sources. Only polymerization and metal-binding regions are shown. *Arabidopsis thaliana and Arachis hypogaea* sequences are used as the standards and highlighted in yellow; the $Mg^{2+}$-binding site is highlighted in light green; the $Zn^{2+}$-binding site is highlighted in orange. The plant system uses the catalytic amino acid R/K as in most of RNAPs, but the template-binding pair is different from the yeast and higher fungi; most of them use an uncommon type of template-binding pair, viz. a –CG- pair and only six of them use a regular –FG- pair (marked in red). Interestingly, all the six of them belong to the family Fabaceae (Leguminosae). However, all of the plant RNAP I template-binding pairs are followed by an invariant H, as in yeast and higher fungi. The $Mg^{2+}$-binding 3 invariant Ds are completely conserved in all (Fig. 6).

CLUSTAL O (1.2.4) MSA of the elongation subunits A1 of RNAP I from yeasts
(Only the polymerization active site and $Mg^{2+}$ binding sites are shown here)



```
//
tr|A0A1E4TEP2|A0A1E4TEP2_9ASCO      SFVGKKVYRHIRNHDVVIMNRQPTLHKASMMGHLVRVLPGEKTLRLHYANTGAYNADFDG      603
tr|A0A7D9CW42|A0A7D9CW42_DEKBR      NAVNKQVFRHIRNGDVVIMNRQPTLHKASMMGHRVRVLPNEKTIRIHYANGGPYNCDYDG      617
tr|A0A448YM75|A0A448YM75_BRENA      SAVNKQVFRHIRNGDVVIMNRQPTLHKASMMGHRVRVLPNEKTLRIHYANGGPYNCDYDG      597
tr|A0A1E5RDD6|A0A1E5RDD6_HANUV      QTLSKKVYRHINNDDIVIMNRQPTLHKASMMGHKVRVLPTEKTLRLHYANTGAYNADFDG      609
tr|A0A1E5R6P8|A0A1E5R6P8_9ASCO      QTLSKKVYRHINNDDIVIMNRQPTLHKASMMGHKVRVLPTEKTLRLHYANTGAYNADFDG      609
tr|A0A1L0CI48|A0A1L0CI48_9ASCO      QTLSKKVYRHINNDDIVIMNRQPTLHKASMMGHKVRVLPTEKTLRLHYANTGAYNADFDG      608
tr|A0A1E5RFI1|A0A1E5RFI1_9ASCO      HTLNKKVYRHIKNNDIVLMNRQPTLHKASMMGHKVRVLPSEKTLRLHYANTGAYNADFDG      635
tr|A0A0X8HTD0|A0A0X8HTD0_9SACH      HALNKKVYRHIKNRDIVIMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      615
tr|A0A1G4MCH6|A0A1G4MCH6_LACFM      HCLNKKVYRHIKNRDVVIMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      615
tr|A0A1G4J4H9|A0A1G4J4H9_9SACH      HSLNKKVYRHIKNRDIVIMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      625
tr|A0A1G4IY91|A0A1G4IY91_9SACH      HSLNKKVYRHIKNRDIVIMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      627
tr|A0A4C2E083|A0A4C2E083_9SACH      HTLNKKVYRHIKNRDIVIMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      631
tr|A0A1Q3ADG7|A0A1Q3ADG7_ZYGRO      HTLNKKVFRHIKNRDIVIMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      631
tr|I2H4E5|I2H4E5_TETBL              HVLNKKVYRHIKNRDIVIMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      632
tr|A0A7G3ZKQ7|A0A7G3ZKQ7_9SACH      HPLNKKVYRHIKDRDIVLMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      636
tr|G9A065|G9A065_TORDC              HPLNKKVYRHIKNRDVVLMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      629
tr|G8BWN2|G8BWN2_TETPH              HTLNKKVYRHIKNRDVVIMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      624
tr|A0A1X7R740|A0A1X7R740_9SACH      HTINKKVYRHIKNRDIVIMNRQPTLHKASMMGHKVRVLPGEKTLRMHYANTGAYNADFDG      632
tr|H2B1A9|H2B1A9_KAZAF              HTLNKKVYRHIKNRDIVIMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      630
tr|G0WET3|G0WET3_NAUDC              HTLNKKVYRHIKNRDIVIMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      633
tr|J8LHJ7|J8LHJ7_SACAR              HTLNKKVYRHIKNRDVVLMNRQPTLHKASMMGHKVRVLPNEKTLRLHYANTGAYNADFDG      632
sp|P10964|RPA1_YEAST                HTLNKKVYRHIKNRDVVLMNRQPTLHKASMMGHKVRVLPNEKTLRLHYANTGAYNADFDG      630
tr|A0A0L8VH35|A0A0L8VH35_9SACH      HTLNKKVYRHIKNRDVVLMNRQPTLHKASMMGHKVRVLPNEKTLRLHYANTGAYNADFDG      630
tr|A0A6C1EA14|A0A6C1EA14_SACPS      HTLNKKVYRHIKNRDVVLMNRQPTLHKASMMGHKVRVLPNEKTLRLHYANTGAYNADFDG      630
tr|Q6C1S4|Q6C1S4_YARLI              GFTNKKVFRHIRNNDVVLMNRQPTLHKASMMGHRVRVLPKEKTLRLHYANTGAYNADFDG      594
tr|A0A7H0KV41|A0A7H0KV41_YARLL      GFTNKKVFRHIRNNDVVLMNRQPTLHKASMMGHRVRVLPKEKTLRLHYANTGAYNADFDG      594
tr|A0A1Z8JTJ3|A0A1Z8JTJ3_PICKU      NVLNKKVYRHIKNNDVVLMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      584
tr|W6MV62|W6MV62_9ASCO              SVVNKKVYRHIKNNDVVLMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      584
tr|A0A4P9ZE88|A0A4P9ZE88_9ASCO      EVVNKKVYRHIRNKDVVLMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      605
tr|A0A512UBI4|A0A512UBI4_9ASCO      SVVNKKVYRHIKNKDVVVMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      609
tr|G3AQ09|G3AQ09_SPAPN              SVVNKKVYRHIKNKDVVIMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      598
tr|M3HQL8|M3HQL8_CANMX              SVVGKKVYRHIRNKDVVIMNRQPTLHKASMMGHKVRVLPGEKTLRLHYANTGAYNADFDG      602
                                       .*:*:***.: *:*:*************** ***** ***:*:**** * **.*:**

tr|A0A1E4TEP2|A0A1E4TEP2_9ASCO      DEMNMHFPQNENARIEAMTLANTDSQYITPTSGKPLRGLIQDHISAGVWLTSKDTFFTRE      663
tr|A0A7D9CW42|A0A7D9CW42_DEKBR      DEMNLHFPQGQLARAESYTLASTDSQYLTPTSGSPVRGLIQDHLSTGVWMTSKNTFFTRE      677
tr|A0A448YM75|A0A448YM75_BRENA      DEMNLHFPQGELARAEAYNLANTDSQYLTPTSGAPVRGLIQDHLSAGVWMTNKDTFFTRE      657
tr|A0A1E5RDD6|A0A1E5RDD6_HANUV      DEMNMHFPQNENARAEALHLANTNSQYLTPTSGAPVRGLIQDHISAGVWLTNKDSFFNRE      669
tr|A0A1E5R6P8|A0A1E5R6P8_9ASCO      DEMNMHFPQNENARAEALHLANTNSQYLTPTSGSPVRGLIQDHISAGVWLTNKDSFFNRE      669
tr|A0A1L0CI48|A0A1L0CI48_9ASCO      DEMNMHFPQNENARAEALHLANTNSQYLTPTSGSPVRGLIQDHISAGVWLTNKDSFFNRE      668
tr|A0A1E5RFI1|A0A1E5RFI1_9ASCO      DEMNMHFPQNENARAEAFNLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTSKDSFFTRE      695
tr|A0A0X8HTD0|A0A0X8HTD0_9SACH      DEMNMHFPQNENARAEAFNLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTNKDSFFTRE      675
tr|A0A1G4MCH6|A0A1G4MCH6_LACFM      DEMNMHFPQNENARAEALNLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTSKDSFFTRE      675
tr|A0A1G4J4H9|A0A1G4J4H9_9SACH      DEMNMHFPQNENARAEAFNLANTDSQYLTPTSGAPVRGLIQDHISAGVWLTSKDSFFTRE      685
tr|A0A1G4IY91|A0A1G4IY91_9SACH      DEMNMHFPQNENARAEAFNLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTSKDSFFTRE      687
tr|A0A4C2E083|A0A4C2E083_9SACH      DEMNMHFPQNENARSEAFNLANTDSQYLTPTSGSPLRGLIQDHISAGVWLTNKDSFFTRD      691
tr|A0A1Q3ADG7|A0A1Q3ADG7_ZYGRO      DEMNMHFPQNENARSEAFNLANTDSQYLTPTSGSPLRGLIQDHISAGVWLTNKDSFFTRD      691
tr|I2H4E5|I2H4E5_TETBL              DEMNMHFPQNENARAESLFLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTNKDSFFTRE      692
tr|A0A7G3ZKQ7|A0A7G3ZKQ7_9SACH      DEMNMHFPQNENARAEAFNLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTSKDSFFTRE      696
tr|G9A065|G9A065_TORDC              DEMNMHFPQNENARAEAFNLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTSKDSFFTRE      689
tr|G8BWN2|G8BWN2_TETPH              DEMNMHFPQNENARAEAFNLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTSKDSYFTRE      684
tr|A0A1X7R740|A0A1X7R740_9SACH      DEMNMHFPQNENARAEAFNLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTNKDSFFTRE      692
tr|H2B1A9|H2B1A9_KAZAF              DEMNMHFPQNENARAEAFNLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTNKDSFFTRE      690
tr|G0WET3|G0WET3_NAUDC              DEMNMHFPQNENARAEAFNLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTSKDSFFTRE      693
tr|J8LHJ7|J8LHJ7_SACAR              DEMNMHFPQNENARAEALNLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTSKDSFFTRE      692
sp|P10964|RPA1_YEAST                DEMNMHFPQNENARAEALNLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTSKDSFFTRE      690
tr|A0A0L8VH35|A0A0L8VH35_9SACH      DEMNMHFPQNENARAEALNLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTSKDSFFTRE      690
tr|A0A6C1EA14|A0A6C1EA14_SACPS      DEMNMHFPQNENARAEALNLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTSKDSFFTRE      690
tr|Q6C1S4|Q6C1S4_YARLI              DEMNMHFPQNENARAEALLLANTDSQYLTPTSGKPLRGLIQDHISAGVLLTSKDTFFTRQ      654
tr|A0A7H0KV41|A0A7H0KV41_YARLL      DEMNMHFPQNENARAEALLLANTDSQYLTPTSGKPLRGLIQDHISAGVLLTSKDSFFTRQ      654
tr|A0A1Z8JTJ3|A0A1Z8JTJ3_PICKU      DEMNMHFPQNENARAEAFNLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTNKDTFFTRE      644
tr|W6MV62|W6MV62_9ASCO              DEMNMHFPQNENAKAEALNLANTDSQYLTPTSGSPVRGLIQDHISAGVWLTSKDSFFTRE      644
tr|A0A4P9ZE88|A0A4P9ZE88_9ASCO      DEMNMHFPQNETARAEALNLANTDSQYLTPTSGAPLRGLIQDHISAGVWLTSKDSFFTRE      665
tr|A0A512UBI4|A0A512UBI4_9ASCO      DEMNMHFPQNENAKAEALNLANTDSQYLTPTSGAPLRGLIQDHISAGVWLTNKDTFFNRE      669
tr|G3AQ09|G3AQ09_SPAPN              DEMNMHFPQNENAKAEALNLANTDNQYLTPTSGSPLRGLIQDHISAGVWLTSKDTFFTRE      658
tr|M3HQL8|M3HQL8_CANMX              DEMNMHFPQNENARAEALNLANTDSQYLTPTSGSPLRGLIQDHISAGVWLTSKDSFFNRE      662
                                    ****:****.* *: *:  **.*:.**:***** *:*******:*:** :*.*:::*.*:
```

//End of the I elongation subunits A1 of RNAP I

```
tr|A0A1E4TEP2|A0A1E4TEP2_9ASCO      CKVGTGSFDIFTPA----------      1654
tr|A0A7D9CW42|A0A7D9CW42_DEKBR      NGSGTGSFDVLAKMPVAQN-----      1620
tr|A0A448YM75|A0A448YM75_BRENA      NKIGTGSFDVLAKMQVKS------      1635
tr|A0A1E5RDD6|A0A1E5RDD6_HANUV      SNVGTGLFDIMAK-----------      1648
tr|A0A1E5R6P8|A0A1E5R6P8_9ASCO      SNVGTGLFDIMAK-----------      1647
tr|A0A1L0CI48|A0A1L0CI48_9ASCO      SNVGTGLFDIMAK-----------      1646
tr|A0A1E5RFI1|A0A1E5RFI1_9ASCO      SGVGTGSFDLLTKSL---------      1672
tr|A0A0X8HTD0|A0A0X8HTD0_9SACH      NAVGTGAFDVLTAVPNAA------      1644
tr|A0A1G4MCH6|A0A1G4MCH6_LACFM      NNVGTGSFDVLAKVPVAN------      1644
tr|A0A1G4J4H9|A0A1G4J4H9_9SACH      NNVGTGSFDVLAKVPASH------      1653
tr|A0A1G4IY91|A0A1G4IY91_9SACH      NNVGTGSFDVLTKVPASH------      1655
tr|A0A4C2E083|A0A4C2E083_9SACH      NNVGTGAFDVLTKIPNSV------      1666
tr|A0A1Q3ADG7|A0A1Q3ADG7_ZYGRO      ----------------------      1645
```

```
tr|I2H4E5|I2H4E5_TETBL          NNVGTGSFDILAKASNNN------        1659
tr|A0A7G3ZKQ7|A0A7G3ZKQ7_9SACH  SNVGTGSFDVLTRVPNNGR-----        1669
tr|G9A065|G9A065_TORDC          SNVGTGSFDLLTKIPNSGV-----        1665
tr|G8BWN2|G8BWN2_TETPH          NNVGTGAFDVLAKASHAN------        1648
tr|A0A1X7R740|A0A1X7R740_9SACH  NGGGTGSFDVLAKVPARA------        1664
tr|H2B1A9|H2B1A9_KAZAF          NNVGTGSFDVLAKVPNAA------        1662
tr|G0WET3|G0WET3_NAUDC          NNVGTGSFDILAKVPNAV------        1669
tr|J8LHJ7|J8LHJ7_SACAR          NNVGTGSFDVLAKVPNAA------        1667
sp|P10964|RPA1_YEAST            NNVGTGSFDVLAKVPNAA------        1664
tr|A0A0L8VH35|A0A0L8VH35_9SACH  NNVGTGSFDVLAKVPNAA------        1664
tr|A0A6C1EA14|A0A6C1EA14_SACPS  NNVGTGSFDVLAKVPNAA------        1664
tr|Q6C1S4|Q6C1S4_YARLI          SGVGTGSFDVKAPIIEEPEE----        1628
tr|A0A7H0KV41|A0A7H0KV41_YARLL  SGVGTGSFDVKAPIIEEPEE----        1628
tr|A0A1Z8JTJ3|A0A1Z8JTJ3_PICKU  NFAGTGSFDLMAKMPDA-------        1600
tr|W6MV62|W6MV62_9ASCO          SNVGTGKFDVMARMPTTAAY----        1609
tr|A0A4P9ZE88|A0A4P9ZE88_9ASCO  NGMGTGSFDVMTRLFISAQDGVMY        1656
tr|A0A512UBI4|A0A512UBI4_9ASCO  NGVGTGSFDVMAKMDKHSPEAV--        1664
tr|G3AQ09|G3AQ09_SPAPN          SNVGTGSFDVFAQIPKGN------        1639
tr|M3HQL8|M3HQL8_CANMX          SKVGTGSFDIFAQMPKI-------        1653
```

**Fig. 4. MSA of the elongation subunits A1 of RNAPs I from different yeasts**

*A0A1E4TEP2_9ASCO Tortispora caseinolytica*
*A0A448YM75_BRENA Brettanomyces naardenensis*
*A0A1E5R6P8_9ASCO Hanseniaspora opuntiae*
*A0A1E5RFI1_9ASCO Hanseniaspora osmophila*
*A0A1G4MCH6_LACFM Lachancea fermentati*
*A0A1G4IY91_9SACH Lachancea* sp.
*A0A1Q3ADG7_ZYGRO Zygosaccharomyces rouxii*
*A0A7G3ZKQ7_9SACH Torulaspora globosa*
*G8BWN2_TETPH Tetrapisispora phaffii*
*H2B1A9_KAZAF Kazachstania Africana*
*J8LHJ7_SACAR Saccharomyces arboricola*
*A0A0L8VH35_9SACH Saccharomyces boulardii*
*Q6C1S4_YARLI Yarrowia lipolytica*
*A0A1Z8JTJ3_PICKU Pichia kudriavzevii*
*A0A4P9ZE88_9ASCO Metschnikowia bicuspidate*
*G3AQ09_SPAPN Spathaspora passalidarum*
*A0A7D9CW42_DEKBR Dekkera bruxellensis*
*A0A1E5RDD6_HANUV Hanseniaspora uvarum*
*A0A1L0CI48_9ASCO Hanseniaspora guilliermondii*
*A0A0X8HTD0_9SACH Eremothecium sinecaudum*
*A0A1G4J4H9_9SACH Lachancea nothofagi*
*A0A4C2E083_9SACH Zygosaccharomyces mellis*
*I2H4E5_TETBL Tetrapisispora blattae*
*G9A065_TORDC Torulaspora delbrueckii*
*A0A1X7R740_9SACH Kazachstania saulgeensis*
*G0WET3_NAUDC Naumovozyma dairenensis*
*P10964|RPA1_YEAST Saccharomyces cerevisiae*
*A0A6C1EA14_SACPS Saccharomyces pastorianus*
*A0A7H0KV41_YARLL Yarrowia lipolytica*
*W6MV62_9ASCO Kuraishia capsulate*
*A0A512UBI4_9ASCO Metschnikowia* sp.
*M3HQL8_CANMX Candida albicans maltose*

CLUSTAL O (1.2.4) MSA of the elongation subunits A1 of RNAP I from higher fungi

```
tr|B6QC93|B6QC93_TALMQ          MASFARPVASELASVDFSVYSSEDIKKISVKRIFNTPSLDSLHNPIPHSLYDPALGAWGD       60
tr|A0A478EDC7|A0A478EDC7_9EURO  MSSFARPVASELASVDFSVYTSEDIKKISVKRIFNTPSLDSLHNPIPNSLYDPALGAWGD       60
tr|W6QED6|W6QED6_PENRF          MATFARPVASTINGVDFNVYSEEEIKALSVKRIHNTPTLDSFNNPVPGGLQDPAMGAWGD       60
tr|A0A167Y3R7|A0A167Y3R7_PENCH  MATFARPVASTINGVDFNVYSEEEIKALSVKRIHNTPTLDSFNNPVPGGLQDPAMGAWGD       60
tr|K9G5R8|K9G5R8_PEND1          MATFARPVASTINGVDFNVYSEEEIKALSVKRIHNTPTLDSFNNPVPGGLQDPAMGAWGD       60
tr|A0A0A2I945|A0A0A2I945_PENEN  MATFARPVASTINGVDFNVYSEEEIKALSVKRIHNTPTLDSFNNPVPGGLQDPAMGAWGD       60
tr|C0NL49|C0NL49_AJECG          MASFTRPISSSIEGVDFRVLSNEEIKTISVKLIYNTPTLDSLNNPVPGGLYDPALGAWGD       60
tr|T5BX79|T5BX79_AJEDE          MASFTRPISSSIGEVDFGVLSNDEIKSISVKRIYNTPTLDTLNNPVPGGLYDPALGAWGD       60
tr|A0A318ZLE2|A0A318ZLE2_9EURO  MATFARPVASTIAGIDFGVYTDEDIKALSVKRIQNTPALDSFNNPVPGGLYDPALGAWGD       60
tr|A0A1L9RJK4|A0A1L9RJK4_ASPWE  MATFARPVASSIGGVDFGVYSNEDIKTISVKRIHNTPTLDSFNNPVPGGLYDPALGAWGD       60
tr|A1D226|A1D226_NEOFI          MATFARPVASSISGIEFGVYSDEDIKSISVKRIHNTPTLDSFNNPVPGGLYDPALGAWGD       60
tr|A0A229XZP5|A0A229XZP5_ASPFM  MATFARPVASSISGIEFGVYSDEDIKSISVKRIHNTPTLDSFNNPVPGGLYDPALGAWGD       60
tr|Q4WS60|Q4WS60_ASPFU          MATFARPVASSISGIEFGVYSDEDIKSISVKRIHNTPTLDSFNNPVPGGLYDPALGAWGD       60
tr|A0A0J5Q544|A0A0J5Q544_ASPFM  MATFARPVASSISGIEFGVYSDEDIKSISVKRIHNTPTLDSFNNPVPGGLYDPALGAWGD       60
tr|A0A3R7HU02|A0A3R7HU02_9EURO  MATFARPVASSISGIEFGVYSDEDIKSISVKRIHNTPTLDSFNNPVPGGLYDPALGAWGD       60
tr|A0A397GCI6|A0A397GCI6_9EURO  MATFARPVASSISGIEFGVYSDEDVKSISVKRIHNTPTLDSFNNPVPGGLYDPALGAWGD       60
                                *::*:**::* :  :** * :.:::* :*** * ***:**::**:* .* ***:*****

tr|B6QC93|B6QC93_TALMQ          HVCTTCRASSWSCPGHPGHIELPVPVWNVTFFDQVYRLLRAKCDYCHRLRMPRIEINAFA       120
tr|A0A478EDC7|A0A478EDC7_9EURO  HVCTTCRASSWSCPGHPGHIELPVPVWNVTFFDQVYRLLRAKCDYCHRFRIARIEIHAYT       120
tr|W6QED6|W6QED6_PENRF          HVCTTCRQNSFTCTGHPGHIELPVPFYNVTFFDHIFRLLRAQCVYCLRLQMGRNQVNMYI       120
tr|A0A167Y3R7|A0A167Y3R7_PENCH  HVCTTCRQNSFTCTGHPGHIELPVPFYNVTFFDHIFRLLRAQCVYCLRLQMGRNQVNMYT       120
tr|K9G5R8|K9G5R8_PEND1          HVCTTCRQNSFTCTGHPGHIELPVPFYNVTFFDHIFRLLRAQCVYCLRLQMGRNQVNMYT       120
tr|A0A0A2I945|A0A0A2I945_PENEN  HVCTTCRQNSFTCTGHPGHIELPVPFYNVTFFDHIFRLLRAQCVYCLRLQMGRNQVNMYT       120
tr|C0NL49|C0NL49_AJECG          HICTTCRLNSWSCNGHAGHIELPVHVYNVTFFDQLFRLLRAQCVYCHRFRMSRADIMSYT       120
tr|T5BX79|T5BX79_AJEDE          HVCTTCRLNSWSCNGHAGHIELPVHVYNVTFFDQLFRLLRAQCAYCHRFRMSRADIKSYT       120
tr|A0A318ZLE2|A0A318ZLE2_9EURO  HVCTTCRQSSWSCTGHSGHVELPVRVYNPTYFEQMYRLLKAQCVYCHRFQMSRFEINTYA       120
tr|A0A1L9RJK4|A0A1L9RJK4_ASPWE  HVCTTCRQNSWSCTGHPGHVELPVHVYNVTFFDQLYRLLRAQCVYCHRLQMSRVNVNAYV       120
tr|A1D226|A1D226_NEOFI          HVCTTCRQNSWSCTGHPGHIELPVRVYNVTFFDQLYRLLRAQCVYCHRFQMSRVQINAYV       120
tr|A0A229XZP5|A0A229XZP5_ASPFM  HVCTTCRQNSWSCTGHPGHIELPVRVYNVTFFDQLYRLLRAQCVYCHRFQMARVQINAYV       120
tr|Q4WS60|Q4WS60_ASPFU          HVCTTCRQNSWSCTGHPGHIELPVRVYNVTFFDQLYRLLRAQCVYCHRFQMARVQINAYV       120
tr|A0A0J5Q544|A0A0J5Q544_ASPFM  HVCTTCRQNSWSCTGHPGHIELPVRVYNVTFFDQLYRLLRAQCVYCHRFQMARVQINAYV       120
tr|A0A3R7HU02|A0A3R7HU02_9EURO  HVCTTCRQNSWSCTGHPGHIELPVRVYNVTFFDQLYRLLRAQCVYCHRFQMPRVQINAYV       120
tr|A0A397GCI6|A0A397GCI6_9EURO  HVCTTCRQNSWSCTGHPGHIELPVRVYNVTFFDQLYRLLRAQCVYCHRFQMPRVQINAYV       120
                                *:**:**** .*::* ** **:**** .:* *:*:::*:**:*:* ** *::: * ::  :
```

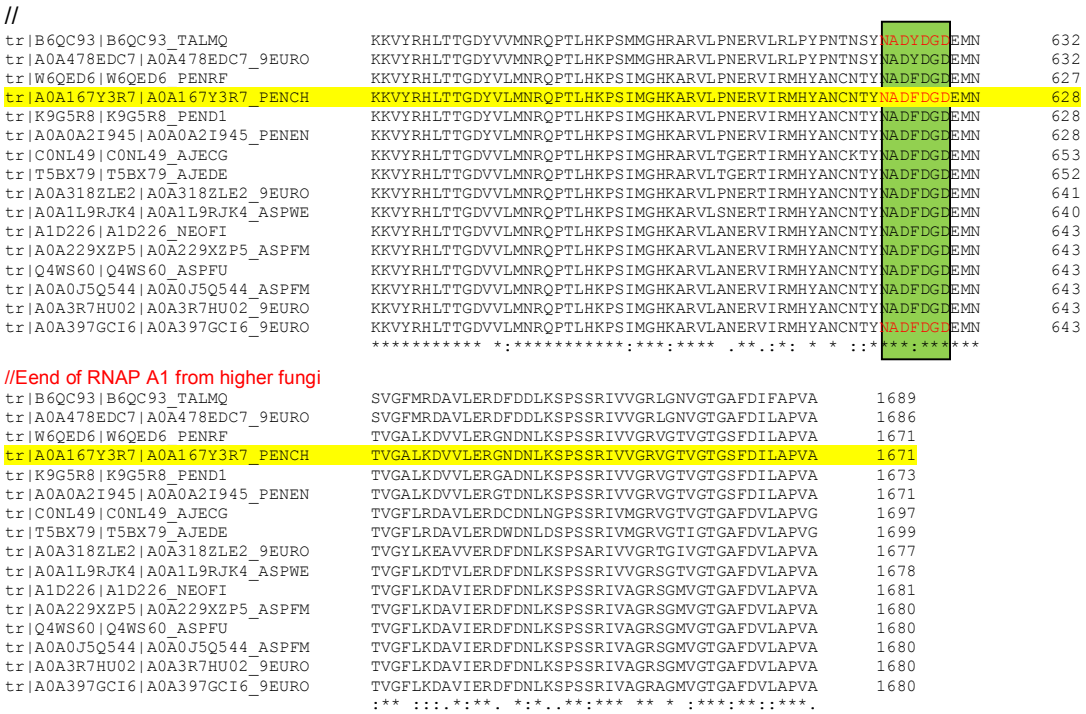
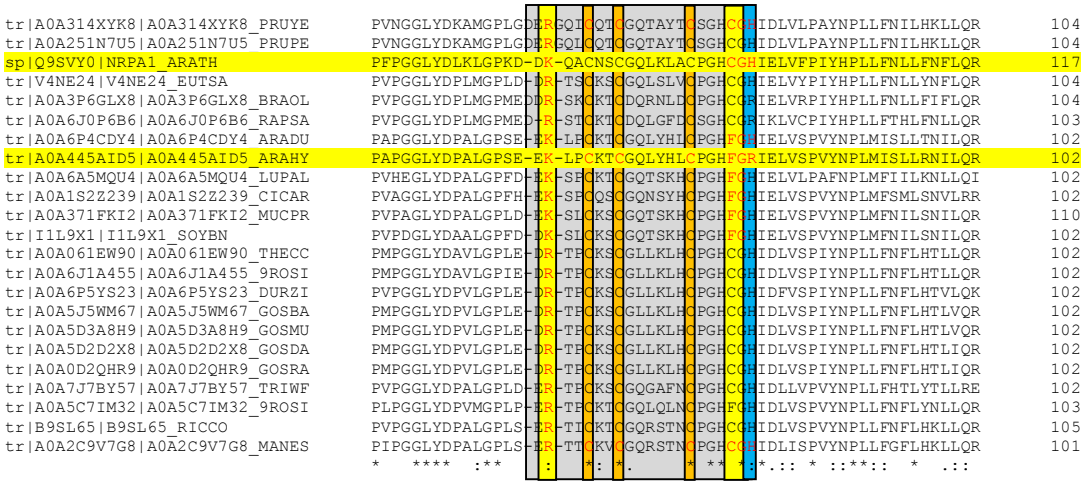



**Fig. 5. MSA of elongation subunits A1 of MSU RNAPs I from higher fungi**

*B6QC93_TALMQ Talaromyces marneffei*
*C0NL49_AJECG Ajellomyces capsulatus*
*W6QED6_PENRF Penicillium roqueforti*
*K9G5R8_PEND1 Penicillium digitatum*
*A0A318ZLE2_9EURO Aspergillus saccharolyticus*
*A1D226_NEOFI Neosartorya fischeri*
*Q4WS60_ASPFU Neosartorya fumigate*
*A0A3R7HU02_9EURO Aspergillus turcosus*
*A0A478EDC7_9EURO Talaromyces cellulolyticus*
*T5BX79_AJEDE Blastomyces dermatitidis*
*A0A167Y3R7_PENCH Penicillium chrysogenum*
*A0A0A2I945_PENEN Penicillium expansum*
*A0A1L9RJK4_ASPWE Aspergillus wentii*
*A0A229XZP5_ASPFM Neosartorya fumigate*
*A0A0J5Q544_ASPFM Aspergillus fumigates*
*A0A397GCI6_9EURO Aspergillus thermomutatus*

CLUSTAL O (1.2.4) MSA of the elongation subunits A1 of RNAP I from plant sources (only polymerization and metal bonding regions are shown)



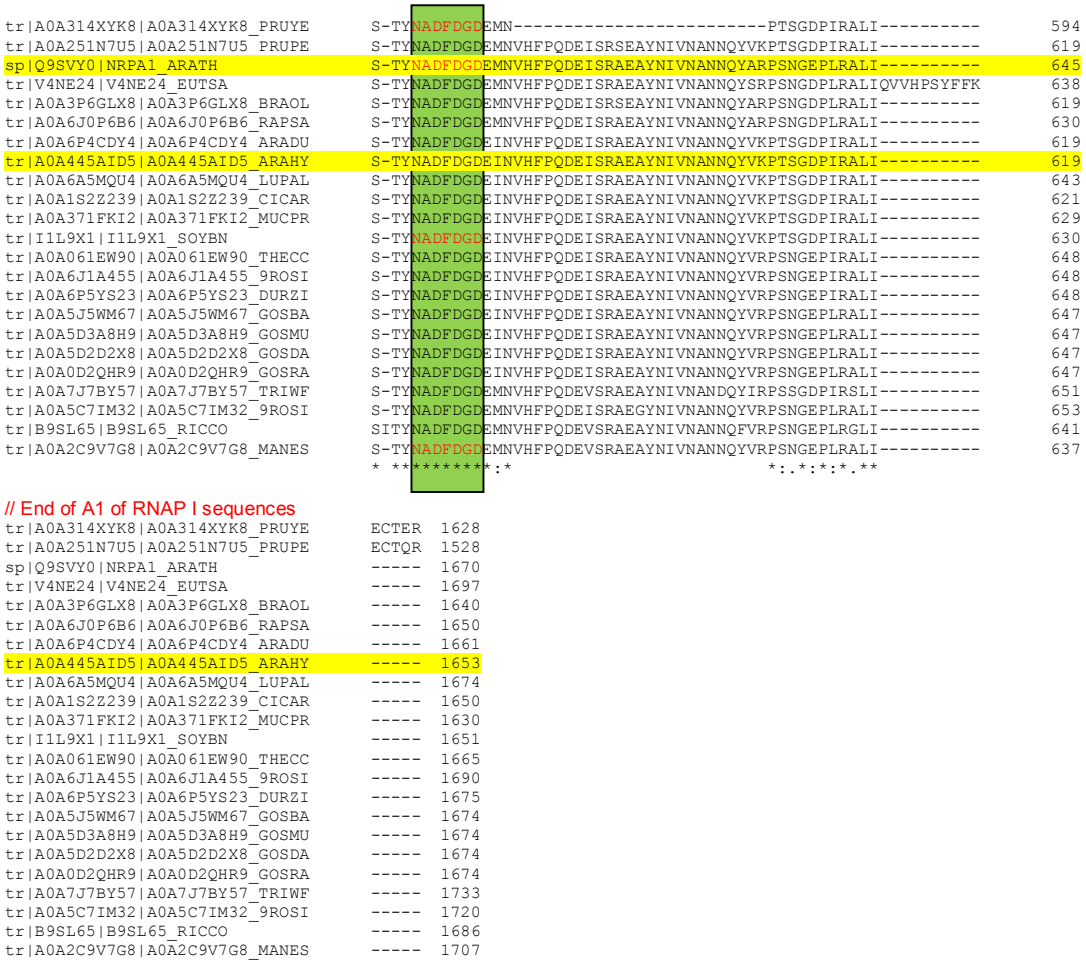
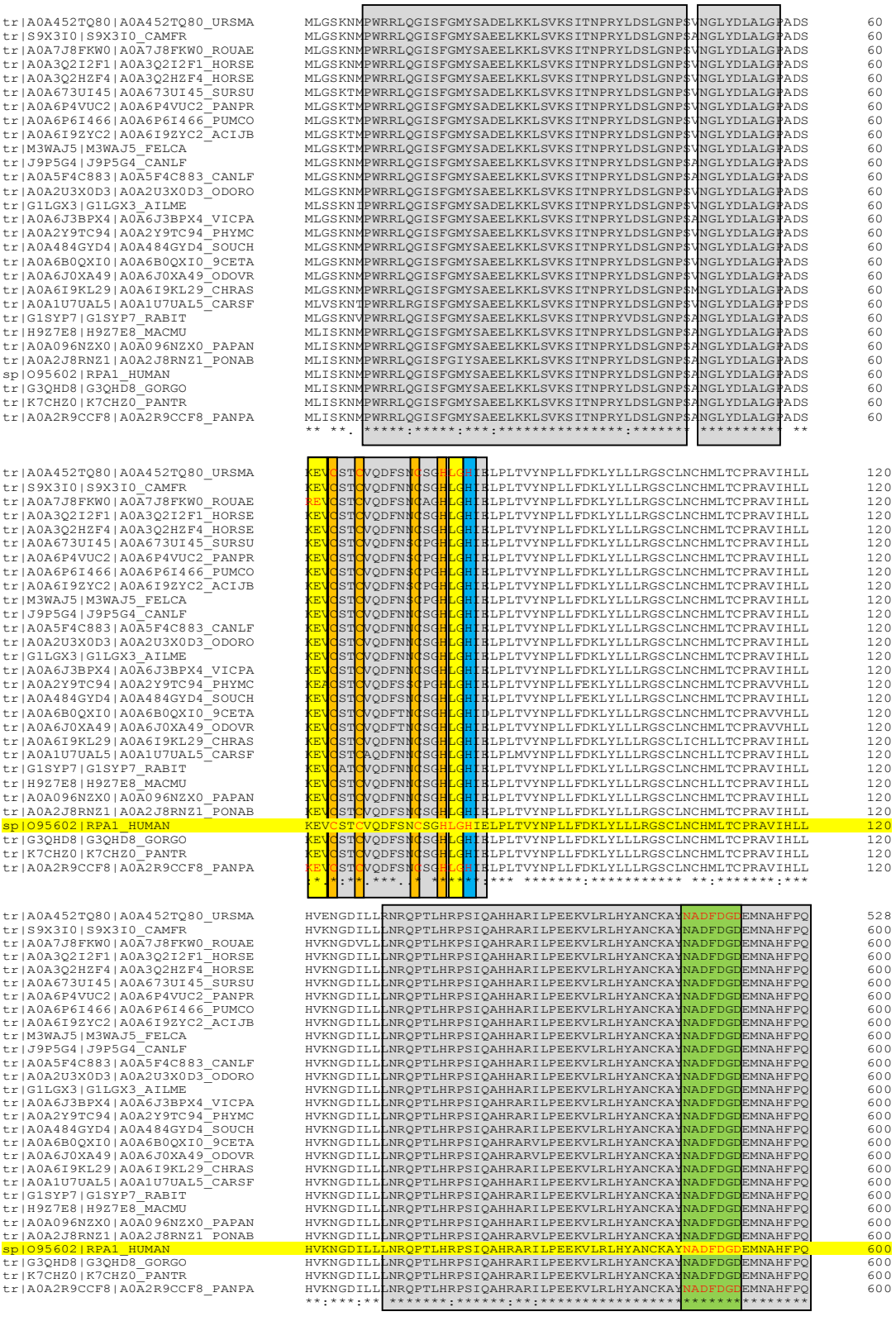



**Fig. 6. MSA of the elongation subunits A1 of RNAPs I from plant sources**

*A0A314XYK8_PRUYE Prunus yedoensis*
*Q9SVY0|NRPA1_ARATH Arabidopsis thaliana*
*A0A3P6GLX8_BRAOL Brassica oleracea*
*A0A6P4CDY4_ARADU Arachis duranensis*
*A0A6A5MQU4_LUPAL Lupinus albus*
*A0A371FKI2_MUCPR Mucuna pruriens*
*A0A061EW90_THECC Theobroma cacao*
*A0A6P5YS23_DURZI Durio zibethinus*
*A0A5D3A8H9_GOSMU Gossypium mustelinum*
*A0A0D2QHR9_GOSRA Gossypium raimondii*
*A0A5C7IM32_9ROSI Acer yangbiense*
*A0A2C9V7G8_MANES Manihot esculenta*
*A0A251N7U5_PRUPE Prunus persica*
*V4NE24_EUTSA Eutrema salsugineum*
*A0A6J0P6B6_RAPSA Raphanus sativus*
*A0A445AID5_ARAHY Arachis hypogaea*
*A0A1S2Z239_CICAR Cicer arietinum*
*I1L9X1_SOYBN Glycine max*
*A0A6J1A455_9ROSI Herrania umbratica*
*A0A5J5WM67_GOSBA Gossypium barbadense*
*A0A5D2D2X8_GOSDA Gossypium darwinii*
*A0A7J7BY57_TRIWF Tripterygium wilfordii*
*B9SL65_RICCO Ricinus communis*

Figure 7 shows the MSA analysis of the elongation subunits A1 of the RNAP I from various animal sources. The human sequence is highlighted in yellow. Remarkably, the N- and C-terminals are almost completely conserved in the elongation subunits of animal RNAPs I. The catalytic and the template-binding pairs are –K/RE- and -LG- in all the sequences. The –LG- pair is followed by an invariant H as found in yeasts, higher fungi and plant sources. The $Mg^{2+}$-binding site is completely conserved in all as found in yeasts, higher fungi and plant sources.

CLUSTAL O (1.2.4) MSA of the elongation subunits A1 of RNAP I from animal sources



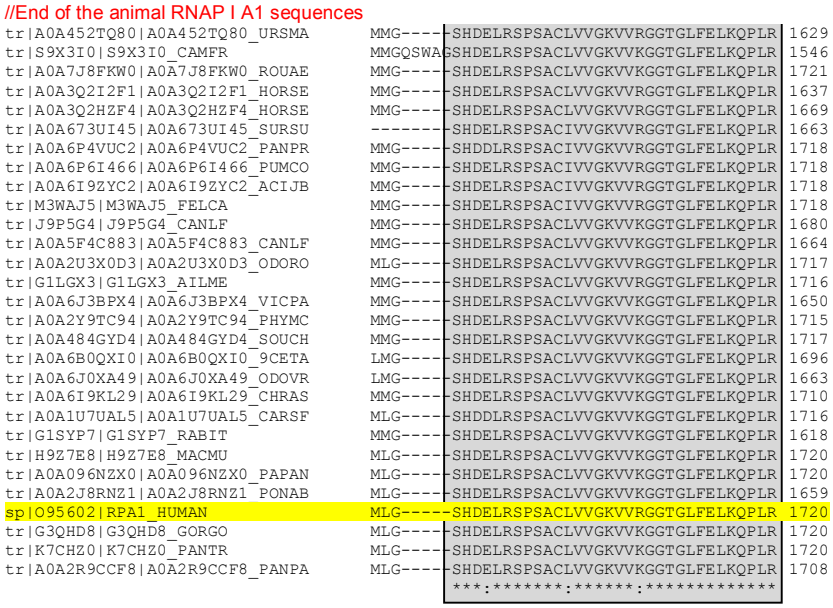



**Fig. 7. MSA of the elongation subunits A1 of RNAP I from animal sources**

*A0A452TQ80_URSMA Ursus maritimus* *S9X3I0_CAMFR Camelus ferus*
*A0A7J8FKW0_ROUAE Rousettus aegyptiacus* *A0A3Q2I2F1_HORSE Equus caballus*
*A0A3Q2HZF4_HORSE Equus caballus,* *A0A673UI45_SURSU Suricata suricatta*
*A0A6P4VUC2_PANPR Panthera pardus* *A0A6P6I466_PUMCO Puma concolor*
*A0A6I9ZYC2_ACIJB Acinonyx jubatus* *M3WAJ5_FELCA Felis catus*
*J9P5G4_CANLF Canis lupus familiaris* *A0A5F4C883_CANLF Canis lupus familiaris*
*A0A2U3X0D3_ODORO Odobenus rosmarus divergens* *G1LGX3_AILME Ailuropoda melanoleuca*
*A0A6J3BPX4_VICPA Vicugna pacos,* *A0A2Y9TC94_PHYMC Physeter macrocephalus*
*A0A484GYD4_SOUCH Sousa chinensis* *A0A6B0QXI0_9CETA Bos mutus*
*A0A6J0XA49_ODOVR Odocoileus virginianus texanus* *A0A6I9KL29_CHRAS Chrysochloris asiatica*
*A0A1U7UAL5_CARSF Carlito syrichta* *G1SYP7_RABIT Oryctolagus cuniculus*
*H9Z7E8_MACMU Macaca mulatta* *A0A096NZX0_PAPAN Papio Anubis*
*A0A2J8RNZ1_PONAB Pongo abelii* *O95602|RPA1_HUMAN Homo sapiens*
*G3QHD8_GORGO Gorilla gorilla gorilla* *K7CHZ0_PANTR Pan troglodytes*
*A0A2R9CCF8_PANPA Pan paniscus*

Therefore, it is clear from the analysis that all of them use a basic amino acid (K/H/R) as the catalytic amino acid, but each group use different types of template-binding pairs as –LG- in animals, -CG- in plants –TG- in higher fungi and -QG- in yeasts.

## 4.2 PR Function in Eukaryotic MSU RNAP II

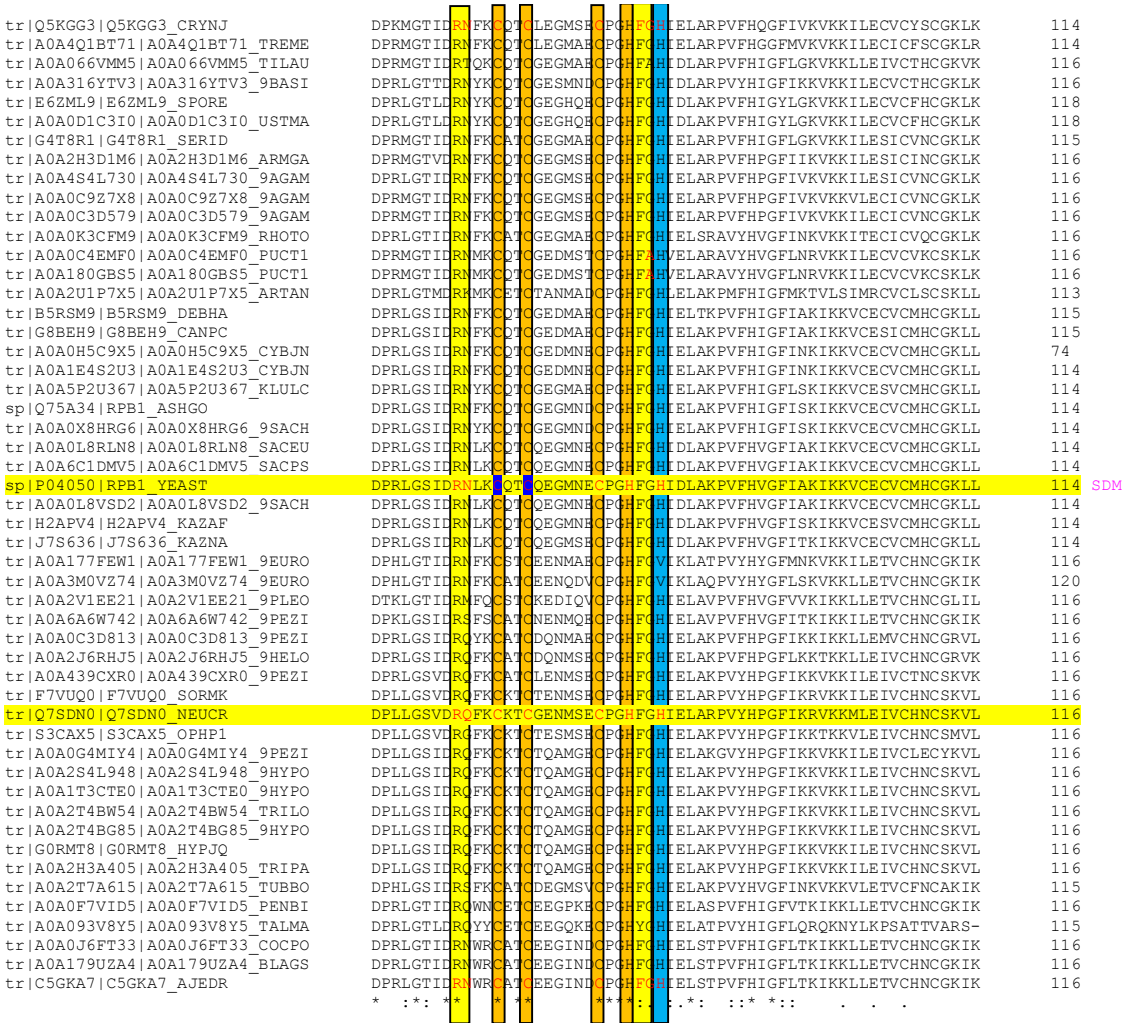
Among the three major MSU RNAPs of eukaryotes, only the RNAP II involves in the transcription of protein-encoding genes (~25,000 genes) resulting in mRNAs. Besides, it also transcribes genes encoding most of the small nuclear RNAs (snRNAs) and microRNAs. In most organisms the RNAP II is made up of 12 subunits (with a molecular mass of ~550 kDa). However, several other proteins are required for complete activity of RNAP II holoenzyme. Therefore, in eukaryotic organisms, the complete 12-subunit RNAP II is responsible for transcription of all protein-encoding genes and thus, forms the central component of the eukaryotic transcription machinery. The RNAP II is strikingly different from other MSU RNAPs. Firstly, it is the only RNAP inhibited by the fungal toxin α-amanitin at very low concentrations, and therefore, all eukaryotic mRNA synthesis is sensitive to this inhibitor. Secondly, it possesses a unique, highly conserved heptapeptide in their carboxyl terminal domain (CTD) [1].

Figure 8 shows the MSA analysis and conserved motifs in the elongation subunits (Rpb1) of the MSU RNAPs II of yeasts and higher fungi. The *S. cerevisiae* (yeast) subunit with 1733 amino acid residues and *Neurospora crassa* (higher fungus) subunit with 1761 amino acid residues are highlighted in yellow. The N-terminal regions are not conserved among them, but the CTD

heptapeptide is highly conserved in all. However, all of them use the catalytic amino acid R. Strikingly the 'template-binding' pair is invariably an –FG- in all, except in few cases where they use -FA- pair. Interestingly, the elongation subunits possessed the possible built-in proposed PR active site with three completely conserved Cs within the polymerases active site region,-$^{55}$DPR$^{-6}$LGSIDRNLKCQTCQEGMNECP GHFGHI$^{84}$- (numbering from *S. cerevisiae* Rpb1 subunit).

However, the Rpb2 initiation subunits of the eukaryotic RNAPs II did not show any built-in PR site; -$^{848}$RGLFR$^{-5}$SLFFRSYMDQEKKYGMS$^{869}$- (numbering from *S. cerevisiae* Rpb2 subunit). Therefore, as found in the prokaryotic elongation subunits, the eukaryotic elongation subunits also possess the PR active site which is integrated within the polymerase active site region (Fig. 8). Unlike its bacterial counterparts, the catalytic region of the elongation subunits of eukaryotes is placed very close to the N-terminal region (within ~100 amino acids) [1]. This Zn-binding site in elongation subunit β' was confirmed by X-ray crystallographic analysis of the MSU RNAP from the thermophilic bacterium, *Thermus aquaticus,* [12]. They found that the $Zn^{2+}$ binds to the 3 invariant Cs located in the catalytic region and also suggested a possible role in the PR activity during elongation. The Zn-binding pattern in the eukaryotic elongation subunits (-$CX_2CX_6C$-) is different from the pattern (-$CX_6CX_2C$-) found in the elongation subunits of eubacteria, i.e., the pattern is just reversed in eukaryotes.

CLUSTAL O (1.2.4) MSA of elongation subunits B1 of RNAP II form yeasts and higher fungi



```
//
tr|Q5KGG3|Q5KGG3_CRYNJ              RHLKDGDYVLFNRQPSLHKMSMMSHRVKLMNYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       509
tr|A0A4Q1BT71|A0A4Q1BT71_TREME      RHLKDGDFVLFNRQPSLHKMSMMCHRVKLMNYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       510
tr|A0A066VMM5|A0A066VMM5_TILAU      RHLKDGDFVLFNRQPSLHKMSMMCHRVKLMDFSTFRLNLSVTPPYNADFDGDEMNLHVPQ       514
tr|A0A316YTV3|A0A316YTV3_9BASI      RHLKDGDFVLFNRQPSLHKMSMMCHRVKLMDFSTFRLNLSVTPPYNADFDGDEMNLHVPQ       512
tr|E6ZML9|E6ZML9_SPORE              RHLKDGDYVLFNRQPSLHKMSMMSHRIKLMDYSTFRLNLSVTPPYNADFDGDEMNLHVPQ       506
tr|A0A0D1C3I0|A0A0D1C3I0_USTMA      RHLKDGDYILFNRQPSLHKMSMMSHRIKLMDYSTFRLNLSVTPPYNADFDGDEMNLHVPQ       506
tr|G4T8R1|G4T8R1_SERID              RHLKDGDYVLFNRQPSLHKMSMMSHRVKIMPYSTFRLNLSVTPPYNADFDGDEMNLHVPQ       505
tr|A0A2H3D1M6|A0A2H3D1M6_ARMGA      RHLKDGDYVLFNRQPSLHKMSMMSHRVRLMPYSTFRLNLSVTPPYNADFDGDEMNMHIPQ       504
tr|A0A4S4L730|A0A4S4L730_9AGAM      RHLKDGDFVLFNRQPSLHKMSMMSHRVRLMPYSTFRLNLSVTPPYNADFDGDEMNMHVPQ       504
tr|A0A0C9Z7X8|A0A0C9Z7X8_9AGAM      RHLKDGDFVLFNRQPSLHKMSMMSHRVKLMPYSTFRLNLSVTPPYNADFDGDEMNMHIPQ       504
tr|A0A0C3D579|A0A0C3D579_9AGAM      RHLKDGDFVLFNRQPSLHKMSMMSHRVRLMPYSTFRLNLSVTPPYNADFDGDEMNMHIPQ       504
tr|A0A0K3CFM9|A0A0K3CFM9_RHOTO      RHLKDGDIVLFNRQPSLHKMSMMAHRVKLMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       505
tr|A0A0C4EMF0|A0A0C4EMF0_PUCT1      RHLKDGDYVLFNRQPSLHKMSMMSHRIKLMPYSTFRLNLSVTPPYNADFDGDEMNLHVPQ       512
tr|A0A180GBS5|A0A180GBS5_PUCT1      RHLKDGDYVLFNRQPSLHKMSMMSHRIKLMPYSTFRLNLSVTPPYNADFDGDEMNLHVPQ       512
tr|A0A2U1P7X5|A0A2U1P7X5_ARTAN      RHLIDGDLVLFNRQPSLHKMSIMGHRIKIMPHSTFRLNLSVTSPYNADFDGDEMNMHVPQ       507
tr|B5RSM9|B5RSM9_DEBHA              RHLMDDDPVLFNRQPSLHKMSMMAHRVKVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       498
tr|G8BEH9|G8BEH9_CANPC              RHLMDDDPVLFNRQPSLHKMSMMAHRVKVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       493
tr|A0A0H5C9X5|A0A0H5C9X5_CYBJN      RHLMDEDPVLFNRQPSLHKMSMMCHRVKVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       451
tr|A0A1E4S2U3|A0A1E4S2U3_CYBJN      RHLMDEDPVLFNRQPSLHKMSMMCHRVKVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       491
tr|A0A5P2U367|A0A5P2U367_KLULC      RHITDNDPVLFNRQPSLHKMSMMAHRVKVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       493
sp|Q75A34|RPB1_ASHGO                RHIMDDDPVLFNRQPSLHKMSMMAHRVKVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       493
tr|A0A0X8HRG6|A0A0X8HRG6_9SACH      RHIMDNDPVLFNRQPSLHKMSMMAHRVKVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       493
tr|A0A0L8RLN8|A0A0L8RLN8_SACEU      RHIMDNDPVLFNRQPSLHKMSMMAHRVKVIPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       493
tr|A0A6C1DMV5|A0A6C1DMV5_SACPS      RHIMDNDPVLFNRQPSLHKMSMMAHRVKVIPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       493
sp|P04050|RPB1_YEAST                RHIMDNDPVLFNRQPSLHKMSMMAHRVKVIPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       493
tr|A0A0L8VSD2|A0A0L8VSD2_9SACH      RHIMDNDPVLFNRQPSLHKMSMMAHRVKVIPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       493
tr|H2APV4|H2APV4_KAZAF              RHIMDDDPVLFNRQPSLHKMSMMSHRVKVVPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       493
tr|J7S636|J7S636_KAZNA              RHIMDDDPVLFNRQPSLHKMSMMAHKVKVIPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       492
tr|A0A177FEW1|A0A177FEW1_9EURO      RHIQDGDVILFNRQPSLHKESMMGHRVRVMPYSTFRLNLSVTTPYNADFDGDEMNLHVPQ       502
tr|A0A3M0VZ74|A0A3M0VZ74_9EURO      RHLQDGDIILFNRQPSLHKESMMAHRVRVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       505
tr|A0A2V1EE21|A0A2V1EE21_9PLEO      RHIDDDDVIIFNRQPSLHKESMMGHRIKVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       500
tr|A0A6A6W742|A0A6A6W742_9PEZI      RHIVDGDVIIFNRQPSLHKESMMGHRVKVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       498
tr|A0A0C3D813|A0A0C3D813_9PEZI      RHIIDGDFIIFNRQPSLHKESMMGHRVRVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       506
tr|A0A2J6RHJ5|A0A2J6RHJ5_9HELO      RHIVDGDFIIFNRQPSLHKESMMGHRVKVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       506
tr|A0A439CXR0|A0A439CXR0_9PEZI      RHLQDGDYIIFNRQPSLHKESMMGHRVKVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       504
tr|F7VUQ0|F7VUQ0_SORMK             RHLIDGDYIIFNRQPSLHKESMMGHRVKVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       507
tr|Q7SDN0|Q7SDN0_NEUCR             RHLIDGDYIIFNRQPSLHKESMMGHRVKVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       507
tr|S3CAX5|S3CAX5_OPHP1             RHLIDGDYIIFNRQPSLHKESMMGHRVRVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       509
tr|A0A0G4MIY4|A0A0G4MIY4_9PEZI      RHLMTGDYIIFNRQPSLHKESMMGHRVRVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       494
tr|A0A2S4L948|A0A2S4L948_9HYPO      RHLMDGDYIIFNRQPSLHKESMMGHRVRVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       506
tr|A0A1T3CTE0|A0A1T3CTE0_9HYPO      RHLMDGDYIIFNRQPSLHKESMMGHRVRVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       507
tr|A0A2T4BW54|A0A2T4BW54_TRILO      RHLMDGDYIIFNRQPSLHKESMMGHRVRVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       507
tr|A0A2T4BG85|A0A2T4BG85_9HYPO      RHLMDGDYIIFNRQPSLHKESMMGHRVRVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       507
tr|G0RMT8|G0RMT8_HYPJQ              RHLMDGDYIIFNRQPSLHKESMMGHRVRVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       507
tr|A0A2H3A405|A0A2H3A405_TRIPA      RHLMDGDYIIFNRQPSLHKESMMGHRVRVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       507
tr|A0A2T7A615|A0A2T7A615_TUBBO      RHIVDGDFIIFNRQPSLHKESMMGHRVRVMPYSTFRLNLSVTSPYNADFDGDEMNLHVPQ       506
tr|A0A0F7VID5|A0A0F7VID5_PENBI      RHLDNGDVILFNRQPSLHKESMMGHRVRVMPFSTFRLNLSVTTPYNADFDGDEMNLHVPQ       505
tr|A0A093V8Y5|A0A093V8Y5_TALMA      RHLLDGDVILFNRQPSLHKESMMGHRVRVMPYSTFRLNLSVTTPYNADFDGDEMNMHVPQ       508
tr|A0A0J6FT33|A0A0J6FT33_COCPO      RHLMDGDVILFNRQPSLHKESMMAHRVRVMPYSTFRLNLSVTTPYNADFDGDEMNLHVPQ       511
tr|A0A179UZA4|A0A179UZA4_BLAGS      RHIVDGDVILFNRQPSLHKESMMGHRIRVMPYSTFRLNLSVTTPYNADFDGDEMNLHVPQ       511
tr|C5GKA7|C5GKA7_AJEDR              RHIVDGDVILFNRQPSLHKESMMGHRIRVMPYSTFRLNLSVTTPYNADFDGDEMNLHVPQ       511
                                    **:   * ::********* *:* *::::: .********** **********:*:**

// End of B1 subunits of RNAP II from yeasts and higher fungi
tr|Q5KGG3|Q5KGG3_CRYNJ              GGYGTSPSWKS----------------      1786
tr|A0A4Q1BT71|A0A4Q1BT71_TREME      GGYGTSPSWKG----------------      1807
tr|A0A066VMM5|A0A066VMM5_TILAU      SNFTSKSMYLDRDQ-------------      1796
tr|A0A316YTV3|A0A316YTV3_9BASI      SRFTQSSMWANKR--------------      1771
tr|E6ZML9|E6ZML9_SPORE              SRMSNKPAWQR----------------      1785
tr|A0A0D1C3I0|A0A0D1C3I0_USTMA      SRMSNKPAWQR----------------      1774
tr|G4T8R1|G4T8R1_SERID              TQYKSSPSWE-----------------      1796
tr|A0A2H3D1M6|A0A2H3D1M6_ARMGA      VSYFSGQRWSQSHLSSVSNPIHHVEY      1802
tr|A0A4S4L730|A0A4S4L730_9AGAM      -AYSASPSYD-----------------      1733
tr|A0A0C9Z7X8|A0A0C9Z7X8_9AGAM      -SYSTSPSWE-----------------      1743
tr|A0A0C3D579|A0A0C3D579_9AGAM      --------------------------      1688
tr|A0A0K3CFM9|A0A0K3CFM9_RHOTO      APYSSSASWKR----------------      1766
tr|A0A0C4EMF0|A0A0C4EMF0_PUCT1      YSYSSAPSWSR----------------      1794
tr|A0A180GBS5|A0A180GBS5_PUCT1      YSYSSAPSWSR----------------      1795
tr|A0A2U1P7X5|A0A2U1P7X5_ARTAN      --------------------------      1728
tr|B5RSM9|B5RSM9_DEBHA              --------------------------      1749
tr|G8BEH9|G8BEH9_CANPC              --------------------------      1746
tr|A0A0H5C9X5|A0A0H5C9X5_CYBJN      --------------------------      1715
tr|A0A1E4S2U3|A0A1E4S2U3_CYBJN      --------------------------      1755
tr|A0A5P2U367|A0A5P2U367_KLULC      --------------------------      1713
sp|Q75A34|RPB1_ASHGO                --------------------------      1745
tr|A0A0X8HRG6|A0A0X8HRG6_9SACH      --------------------------      1725
tr|A0A0L8RLN8|A0A0L8RLN8_SACEU      --------------------------      1719
tr|A0A6C1DMV5|A0A6C1DMV5_SACPS      --------------------------      1726
sp|P04050|RPB1_YEAST                --------------------------      1733
tr|A0A0L8VSD2|A0A0L8VSD2_9SACH      --------------------------      1733
```

```
tr|J7S636|J7S636_KAZNA          ------------------------    1721
tr|A0A177FEW1|A0A177FEW1_9EURO  ------------------------    1762
tr|A0A3M0VZ74|A0A3M0VZ74_9EURO  ------------------------    1771
tr|A0A2V1EE21|A0A2V1EE21_9PLEO  ------------------------    1755
tr|A0A6A6W742|A0A6A6W742_9PEZI  ------------------------    1762
tr|A0A0C3D813|A0A0C3D813_9PEZI  ------------------------    1739
tr|A0A2J6RHJ5|A0A2J6RHJ5_9HELO  ------------------------    1737
tr|A0A439CXR0|A0A439CXR0_9PEZI  ------------------------    1737
tr|F7VUQ0|F7VUQ0_SORMK          ------------------------    1761
tr|Q7SDN0|Q7SDN0_NEUCR          ------------------------    1761
tr|S3CAX5|S3CAX5_OPHP1          ------------------------    1757
tr|A0A0G4MIY4|A0A0G4MIY4_9PEZI  ------------------------    1743
tr|A0A2S4L948|A0A2S4L948_9HYPO  ------------------------    1755
tr|A0A1T3CTE0|A0A1T3CTE0_9HYPO  ------------------------    1756
tr|A0A2T4BW54|A0A2T4BW54_TRILO  ------------------------    1561
tr|A0A2T4BG85|A0A2T4BG85_9HYPO  ------------------------    1755
tr|G0RMT8|G0RMT8_HYPJQ          ------------------------    1755
tr|A0A2H3A405|A0A2H3A405_TRIPA  ------------------------    1755
tr|A0A2T7A615|A0A2T7A615_TUBBO  ------------------------    1751
tr|A0A0F7VID5|A0A0F7VID5_PENBI  ------------------------    1733
tr|A0A093V8Y5|A0A093V8Y5_TALMA  ------------------------    1743
tr|A0A0J6FT33|A0A0J6FT33_COCPO  ------------------------    1750
tr|A0A179UZA4|A0A179UZA4_BLAGS  ------------------------    1746
tr|C5GKA7|C5GKA7_AJEDR          ------------------------    1746
```

**Fig. 8. MSA of elongation subunits B1 of RNAPs II from yeasts and higher fungal sources**

*Q5KGG3_CRYNJ Cryptococcus neoformans*
*A0A066VMM5_TILAU Tilletiaria anomala*
*E6ZML9_SPORE Sporisorium reilianum*
*G4T8R1_SERID Serendipita indica*
*A0A4S4L730_9AGAM Bondarzewia mesenterica*
*A0A0C3D579_9AGAM Scleroderma citrinum*
*A0A0C4EMF0_PUCT1 Puccinia triticina*
*A0A2U1P7X5_ARTAN Artemisia annua*
*G8BEH9_CANPC Candida parapsilosis*
*A0A1E4S2U3_CYBJN Cyberlindnera jadinii*
Q75A34|*RPB1_ASHGO Ashbya gossypii*
A0A0L8RLN8_SACEU *Saccharomyces eubayanus*
*P04050|RPB1_YEAST Saccharomyces cerevisiae*
*H2APV4_KAZAF Kazachstania africana*
*A0A177FEW1_9EURO Fonsecaea monophora*
*A0A2V1EE21_9PLEO Periconia macrospinosa*
*A0A0C3D813_9PEZI Oidiodendron maius*
*A0A439CXR0_9PEZI Xylaria grammica*
*Q7SDN0_NEUCR Neurospora crassa*
*A0A0G4MIY4_9PEZI Verticillium longisporum*
*A0A1T3CTE0_9HYPO Trichoderma guizhouense*
*A0A2T4BG85_9HYPO Trichoderma citrinoviride*
*A0A2H3D1M6_ARMGA Armillaria gallica*
*A0A0F7VID5_PENBI ,Penicillium brasilianum*
*A0A0J6FT33_COCPO Coccidioides posadasii*
*C5GKA7_AJEDR Ajellomyces dermatitidis*
*A0A4Q1BT71_TREME Tremella mesenterica*
*A0A316YTV3_9BASI Acaromyces ingoldii*
*A0A0D1C3I0_USTMA Ustilago maydis*
*A0A2H3A405_TRIPA Trichoderma parareesei*
*A0A0C9Z7X8_9AGAM Pisolithus microcarpus*
*A0A0K3CFM9_RHOTO Rhodosporidium toruloides*
*A0A180GBS5_PUCT1 Puccinia triticina*
*B5RSM9_DEBHA Debaryomyces hansenii*
*A0A0H5C9X5_CYBJN Cyberlindnera jadinii*
*A0A5P2U367_KLULC Kluyveromyces lactis*
*A0A0X8HRG6_9SACH Eremothecium sinecaudum*
*A0A6C1DMV5_SACPS Saccharomyces pastorianus*
*A0A0L8VSD2_9SACH Saccharomyces boulardii*
*J7S636_KAZNA ,Kazachstania naganishii*
*A0A3M0VZ74_9EURO Chaetothyriales sp.*
*A0A6A6W742_9PEZI Pseudovirgaria hyperparasitica*
*A0A2J6RHJ5_9HELO Hyaloscypha variabilis*
*F7VUQ0_SORMK ,Sordaria macrospora*
*S3CAX5_OPHP1 Ophiostoma piceae*
*A0A2S4L948_9HYPO Tolypocladium paradoxum*
*A0A2T4BW54_TRILO Trichoderma longibrachiatum*
*G0RMT8_HYPJQ Hypocrea jecorina*
*A0A2T7A615_TUBBO Tuber borchii*
*A0A093V8Y5_TALMA Talaromyces marneffei*
*A0A179UZA4_BLAGS Blastomyces gilchristii*

Further insights into the zinc-binding site in eukaryotes were provided by Crammer [23]. X-ray crystallographic data for the 10-subunit eukaryotic (yeast) RNAP II at 2.8 Å and 3.1 Å resolutions were reported. A 2.8 Å difference Fourier map revealed two metal ions at the active site, one persistently bound and the other possibly exchangeable during RNA synthesis, suggesting a $Zn^{2+}$ in the exchangeable site. The $Mg^{2+}$ (Metal A) is coordinated by amino acid residues D483 and D485 from the completely conserved block showing three invariant Ds, (D481, D483, and D485). Metal B is in the vicinity of metal A, at a distance of 5.8 Å. The distance from metal B to the acidic residues are at 3.0 Å to 4.0 Å, and hence is too great for coordination of $Zn^{2+}$. They also found a $Zn^{2+}$ is bound by residues in the common motif $CX_2CX_nCX_2C/H$ (where X is any amino acid) [23].

Donaldson and Friesen found that a highly purified yeast RNAP II bound to 7 $Zn^{2+}$ atoms by atomic absorption spectroscopy. One of the zinc-binding motifs ($^{67}CX_2CX_6CX_2HX_{26}CX_2C^{110}$) occurs in the Rpb1 elongation subunit of RNAP II and is found to be highly conserved in the largest subunits (elongation subunits) of all three RNA polymerases from a variety of eukaryotes [24].

They also proved by SDM experiments that mutations of the $Zn^{2+}$-coordinating C residues conferred a lethal phenotype (C77→S and H80→Y were unable to support growth at 37° C). Similar results were obtained for the *E. coli* RNAP too. For example, in the double mutant where the first two Cs of the zinc-binding motif were modified to S (C→S) in the β' elongation subunit of *E. coli* RNAP, the enzyme was found to be inactive *in vivo* [25].

Figure 9 shows the MSA analysis of the elongation subunits B1 of RNAP II from various plant sources. *Triticum aestivum (w*heat) and *A. thaliana* standard sequences are highlighted in yellow, the template-binding and catalytic pairs are highlighted in yellow and the Zn-binding conserved Cs is highlighted in orange. The completely conserved $Mg^{2+}$ -binding site is highlighted in light green. No conservation is seen in the N-terminal region among them, but the CTD exhibited the highly conserved heptapeptide. The catalytic amino acid is conserved in all and uses the basic amino acid K with one exception where it uses an R. Strikingly, the template-binding FG pair and the 3 invariant Ds in the $Mg^{2+}$ binding site are completely conserved in all.

Figure 10 shows the MSA analysis of the elongation subunits B1 of RNAP II from various animal and insect sources. The human sequence is highlighted in yellow. The template-binding and catalytic pairs are highlighted in yellow and the Zn-binding conserved Cs is highlighted in orange. The completely conserved $Mg^{2+}$ site is highlighted in light green. All of them use the basic amino acid R unlike in plants where they use a K. The template-binding pair –FG- is also completely conserved in all except in *Drosophila albomicans.* The $Mg^{2+}$-binding site is completely conserved in all with the 3 invariant Ds. The CTD possess the characteristic heptapeptide repeats -YSPTSPT-, a unique feature of these enzymes.

CLUSTAL O (1.2.4) MSA of elongation subunits B1 of RNAP II from form plant sources

```
tr|A0A1Z5RIH1|A0A1Z5RIH1_SORBI      PKAGGLSDPRMGTVDRKIKQQTCMSGMAECPGHFGYLELAKPMFHIGFMKTVLSIMRCVC      106
tr|C5Y387|C5Y387_SORBI              PKAGGLSDPRMGTVDRKIKQQTCMSGMAECPGHFGHLELAKPMFHIGFMKTVLSIMRCVC      106
tr|A0A3B6MZU3|A0A3B6MZU3_WHEAT      PKPGGLSDPRLGTIDRRTKCETCMAGMAECPGHFGHLELAKPMFHIGFIKTVLSIMRCVC      106
tr|A0A565C923|A0A565C923_9BRAS      PKVGGLSDMRLGTIDRKVKCETCMANMAECPGHFGHLELAKPMYHVGFMKTVLSIMRCVC      76
tr|M4D210|M4D210_BRARP              PKVGGLSDTRLGTIDRKVKCETCMANMAECPGHFGHLELAKPMYHVGFMKTVLSIMRCVC      106
tr|A0A6J0KQ09|A0A6J0KQ09_RAPSA      PKVGGLSDTRLGTIDRKVKCETCMANMAECPGHFGHLELAKPMYHVGFMKTVLSIMRCVC      106
tr|A0A1J3F625|A0A1J3F625_NOCCA      PKVGGLSDVRLGTIDRKVKCETCMANMAECPGHFGHLELAKPMYHVGFMKTVLSIMRCVC      106
tr|A0A1J3JR03|A0A1J3JR03_NOCCA      PKVGGLSDVRLGTIDRKVKCETCMANMAECPGHFGHLELAKPMYHVGFMKTVLSIMRCVC      106
sp|P18616|NRPB1_ARATH               PKVGGLSDTRLGTIDRKVKCETCMANMAECPGHFGYLELAKPMYHVGFMKTVLSIMRCVC      106
tr|D7MC36|D7MC36_ARALL              PKVGGLSDIRLGTIDRKVKCETCMANMAECPGHFGHLELAKPMYHVGFMKTVLSIMRCVC      106
tr|V4P2F5|V4P2F5_EUTSA              PKVGGLSDIRLGTIDRKVKCETCMANMAECPGHFGHLELAKPMFHVGFMKTVLSIMRCVC      106
tr|R0GXL7|R0GXL7_9BRAS              PKVGGLSDIRLGTIDRKVKCETCMANMAECPGHFGHLELAKPMYHVGFMKTVLSIMRCVC      106
tr|A0A6J0MG42|A0A6J0MG42_RAPSA      PKVGGLSDVRLGTIDRKVKCETCMANMAECPGHFGHLELAKPMYHVGFMKTVLSIMRCVC      106
tr|A0A398AHR5|A0A398AHR5_BRACM      PKVGGLSDARLGTIDRKVKCETCMANMAECPGHFGHLELAKPMYHVGFMKTVLSIMRCVC      106
tr|A0A0D3A1P5|A0A0D3A1P5_BRAOL      PKVGGLSDARLGTIDRKVKCETCMANMAECPGHFGHLELAKPMYHVGFMKTVLSIMRCVC      106
tr|A0A3N6QGC8|A0A3N6QGC8_BRACR      PKVGGLSDARLGTIDRKVKCETCMANMAECPGHFGHLELAKPMYHVGFMKTVLSIMRCVC      106
tr|A0A0E0LRD6|A0A0E0LRD6_ORYPU      PKPGGLSDPRLGTIDRKIKCETCMAGMAECPGHFGHLELAKPMFHIGFIKTVLSIMRCVC      106
tr|K3YFU7|K3YFU7_SETIT              PKPGGLSDPRLGTIDRKIKCETCMAGMAECPGHFGHLELAKPMFHIGFIKTVLSIMRCVC      106
tr|J3MQH6|J3MQH6_ORYBR              PKPGGLSDPRLGTIDRKIKCETCMAGMAECPGHFGHLELAKPMFHIGFIKTVLSIMRCVC      106
tr|A0A803MAV7|A0A803MAV7_CHEQI      PKPGGLSDMRLGTIDRKIKCDTCTANMAECPGHFGHLELAKPMFHIGFLKTVLSIMRCVC      106
tr|A0A0K9RSK8|A0A0K9RSK8_SPIOL      PKPGGLSDMRLGTIDRKIKCDTCTANMAECPGHFGHLELAKPMFHIGFLKTVLSIMRCVC      106
tr|A0A6P6TFX2|A0A6P6TFX2_COFAR      PKIGGLSDPRLGTIDRKMKCETCMASMAECPGHFGHLELAKPMFHIGFIKTVLSIMRCVC      106
tr|A0A6I9UDD3|A0A6I9UDD3_SESIN      PKPGGLSDPRLGTIDRKMKCETCMANMADCPGHFGHLELAKPMFHIGFMKTVLSILRCVC      106
tr|A0A1S3XM16|A0A1S3XM16_TOBAC      PKPGGLSDPRLGTIDRKMKCETCMANMAECPGHFGHLELAKPMFHIGFMKPVLSILRCVC      106
tr|A0A2G3A9T8|A0A2G3A9T8_CAPAN      PKPGGLSDPRLGTIDRKMKCETCMANMAECPGHFGHLELAKPMFHIGFMKPVLSILRCVC      106
tr|A0A1U8FXJ4|A0A1U8FXJ4_CAPAN      PKPGGLSDPRLGTIDRKMKCETCMANMAECPGHFGHLELAKPMFHIGFMKPVLSILRCVC      106
tr|M0ZVV4|M0ZVV4_SOLTU              PKPGGLSDPRLGTIDRKMKCETCMANMAECPGHFGHLELAKPMFHIGFMKPVLSILRCVC      106
tr|A0A6N2BNI8|A0A6N2BNI8_SOLCI      PKPGGLSDPRLGTIDRKMKCETCMANMAECPGHFGHLELAKPMFHIGFMKPVLSILRCVC      106
tr|A0A3Q7F7G2|A0A3Q7F7G2_SOLLC      PKPGGLSDPRLGTIDRKMKCETCMANMAECPGHFGHLELAKPMFHIGFMKPVLSILRCVC      106
tr|A0A6V7PIB4|A0A6V7PIB4_ANACO      AEARGLSDPRLGTIDRKMKCETCMANMAECPGHFGHLELAKPMFHIGFLKTVLAIMRCVC      120
tr|A0A6I9RZH4|A0A6I9RZH4_ELAGV      PKPGGLSDPRLGTIDRKMKCETCMANMAECPGHFGHLELAKPMFHIGFLKTVLAIMRCVC      106
tr|A0A4Y7I7U1|A0A4Y7I7U1_PAPSO      PKPGGLSDTRLGTIDRKMKCETCTANMAECPGHFGHLELAKPMFHIGFMKIVLSIMRCVC      106
tr|A0A6P9E4T1|A0A6P9E4T1_JUGRE      PKIAGLSDPRLGTIDRKMKCETCTANMAECPGHFGHLELAKPMFHIGFMKTVLSIMRCVC      106
tr|A0A7J7HU38|A0A7J7HU38_CAMSI      PKPGGLSDPRLGTIDRKMKCETCTANMAECPGHFGHLELAKPMFHIGFLKTVLSIMRCVC      106
tr|F6H0D9|F6H0D9_VITVI              PKPGGLSDPRLGTIDRKMKCETCTANMAECPGHFGHLELAKPMFHIGFMKTVLSIMRCVC      106
tr|A0A7J6EME3|A0A7J6EME3_CANSA      PKIAGLSDPRLGTIDRKMKCETCTANMAECPGHFGHLELAKPMFHIGFLKTVLSIMRCVC      106
tr|A0A067G7I3|A0A067G7I3_CITSI      PKPGGLSDPRLGTIDRKMKCETCTANMAECPGHFGHLELAKPMFHIGFMKTVLSIMRSVC      106
tr|A0A5J5TPI9|A0A5J5TPI9_GOSBA      PKVGGLSDPRLGTIDRKIKCETCTANMAECPGHFGHLELAKPMFHIGFMKTVLSIMRCVC      106
tr|A0A061DGE8|A0A061DGE8_THECC      PKVGGLSDPRLGTIDRKMKCETCTANMAECPGHFGHLELAKPMFHIGFMKTVLSIMRCVC      106
tr|A0A6P5ZFF2|A0A6P5ZFF2_DURZI      PKVGGLSDPRLGTIDRKMKCETCTANMAECPGHFGHLELAKPMFHIGFMKTVLSIMRCVC      106
tr|B9SXC0|B9SXC0_RICCO              PKPGGLSDLRLGTIDRKMKCETCTANMAECPGHFGHLELAKPMFHIGFLKTVLSIMRCVC      106
tr|A0A251M2B5|A0A251M2B5_MANES      PKPGGLSDPKLGTIDRKMKCETCTANMAECPGHFGHLELAKPMFHIGFMKTVLSIMRCVC      106
tr|A0A314Z8L7|A0A314Z8L7_PRUYE      PKTAGLSDPRLGTIDRKMKCETCTANMAECPGHFGHLELAKPMFHIGFMKTVLSVMRCVC      106
tr|M5VSC7|M5VSC7_PRUPE              PKTAGLSDPRLGTIDRKMKCETCTANMAECPGHFGHLELAKPMFHIGFMKTVLSIMRCVC      106
tr|A0A6J5XYF9|A0A6J5XYF9_PRUAR      PKTAGLSDPRLGTIDRKMKCETCTANMAECPGHFGHLELAKPMFHIGFMKTVLSVMRCVC      106
tr|A0A6J1L0Z7|A0A6J1L0Z7_CUCMA      PKVAGLSDPRLGTIDRKIKCETCTANMAECPGHFGHLELAKPMFHIGFMKTVLTIMRSVC      106
tr|A0A0A0L655|A0A0A0L655_CUCSA      PKVAGLSDPRLGTIDRKMKCETCTANMAECPGHFGYLELAKPMFHIGFMKTVLTIMRSVC      106
                                     :  **** ::** .:: **:** ..:*:******* *******:*:**:* **:::*.**
```

```
//
tr|A0A1Z5RIH1|A0A1Z5RIH1_SORBI      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNKPVMGIVQDTLLGCRKITK      549
tr|C5Y387|C5Y387_SORBI              PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNKPVMGIVQDTLLGCRKITK      549
tr|A0A3B6MZU3|A0A3B6MZU3_WHEAT      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANKPVMGIVQDTLLGCRKITK      551
tr|A0A565C923|A0A565C923_9BRAS      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKITK      520
tr|M4D210|M4D210_BRARP              PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKITK      536
tr|A0A6J0KQ09|A0A6J0KQ09_RAPSA      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKITK      550
tr|A0A1J3F625|A0A1J3F625_NOCCA      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKITK      550
tr|A0A1J3JR03|A0A1J3JR03_NOCCA      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKITK      550
sp|P18616|NRPB1_ARATH               PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKITK      550
tr|D7MC36|D7MC36_ARALL              PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKITK      551
tr|V4P2F5|V4P2F5_EUTSA              PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKITK      550
tr|R0GXL7|R0GXL7_9BRAS              PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKITK      550
tr|A0A6J0MG42|A0A6J0MG42_RAPSA      PYNADFDGDEMNMHVPQSYETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKITK      550
tr|A0A398AHR5|A0A398AHR5_BRACM      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKITK      550
tr|A0A0D3A1P5|A0A0D3A1P5_BRAOL      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKITK      550
tr|A0A3N6QGC8|A0A3N6QGC8_BRACR      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKITK      546
tr|A0A0E0LRD6|A0A0E0LRD6_ORYPU      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      549
tr|K3YFU7|K3YFU7_SETIT              PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      549
tr|J3MQH6|J3MQH6_ORYBR              PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      549
tr|A0A803MAV7|A0A803MAV7_CHEQI      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKVTK      551
tr|A0A0K9RSK8|A0A0K9RSK8_SPIOL      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANKPVMGIVQDTLLGCRKVTK      549
tr|A0A6P6TFX2|A0A6P6TFX2_COFAR      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      549
tr|A0A6I9UDD3|A0A6I9UDD3_SESIN      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      549
tr|A0A1S3XM16|A0A1S3XM16_TOBAC      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKITK      549
tr|A0A2G3A9T8|A0A2G3A9T8_CAPAN      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKVTK      549
tr|A0A1U8FXJ4|A0A1U8FXJ4_CAPAN      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKVTK      549
tr|M0ZVV4|M0ZVV4_SOLTU              PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKVTK      549
tr|A0A6N2BNI8|A0A6N2BNI8_SOLCI      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKVTK      549
tr|A0A3Q7F7G2|A0A3Q7F7G2_SOLLC      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKVTK      549
tr|A0A6V7PIB4|A0A6V7PIB4_ANACO      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      564
tr|A0A6I9RZH4|A0A6I9RZH4_ELAGV      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      550
tr|A0A4Y7I7U1|A0A4Y7I7U1_PAPSO      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      549
tr|A0A6P9E4T1|A0A6P9E4T1_JUGRE      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      549
tr|A0A7J7HU38|A0A7J7HU38_CAMSI      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGCRKITK      567
tr|F6H0D9|F6H0D9_VITVI              PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      553
tr|A0A7J6EME3|A0A7J6EME3_CANSA      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      504
tr|A0A067G7I3|A0A067G7I3_CITSI      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      539
tr|A0A5J5TPI9|A0A5J5TPI9_GOSBA      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      549
tr|A0A061DGE8|A0A061DGE8_THECC      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      549
tr|A0A6P5ZFF2|A0A6P5ZFF2_DURZI      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      549
tr|B9SXC0|B9SXC0_RICCO              PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      549
tr|A0A251M2B5|A0A251M2B5_MANES      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      549
tr|A0A314Z8L7|A0A314Z8L7_PRUYE      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      549
tr|M5VSC7|M5VSC7_PRUPE              PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      549
tr|A0A6J5XYF9|A0A6J5XYF9_PRUAR      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      549
tr|A0A6J1L0Z7|A0A6J1L0Z7_CUCMA      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDSLLGCRKITK      549
tr|A0A0A0L655|A0A0A0L655_CUCSA      PYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQSNRPVMGIVQDTLLGCRKITK      549
                                    *****************: ********************:*:********:******:**
```

// End of elongation subunits B1 of RNAP II from plants

```
tr|A0A1Z5RIH1|A0A1Z5RIH1_SORBI      ----------------------------------------------------      1633
tr|C5Y387|C5Y387_SORBI              SPNFSPNRSDSPTAADYSPSSTGQAIEKDGETSH------------------      1874
tr|A0A3B6MZU3|A0A3B6MZU3_WHEAT      SPNYSPSASYSPSSTGPQTTDMDD----------ETAT---------------      1830
tr|A0A565C923|A0A565C923_9BRAS      SPDYSPSAGYSPTLPGYSPSSTGQYTPHEGDKNDKTGKD-ASKDDKSKP----      1806
tr|M4D210|M4D210_BRARP              SPDYSPSAGYSPTLPGYSPSSTGQYTPHEGDKNDKTGKD-AGTDDKSKP----      1824
tr|A0A6J0KQ09|A0A6J0KQ09_RAPSA      SPDYSPSAGYSPTLPGYSPSSTGQYTPHEGDKNDKTGKD-GSKDDKSNP----      1838
tr|A0A1J3F625|A0A1J3F625_NOCCA      SPDYSPSAGYSPTLPGYSPSSTGQFTPHEGDKNDQTGK-DVSKDDKTNP----      1838
tr|A0A1J3JR03|A0A1J3JR03_NOCCA      SPDYSPSAGYSPTLPGYSPSSTGQFTPHEGDKNDQTGK-DVSKDDKTNP----      1838
sp|P18616|NRPB1_ARATH               SPDYSPSAGYSPTLPGYSPSSTGQYTPHEGDKKDKTGKKDASKDDKGNP----      1839
tr|D7MC36|D7MC36_ARALL              SPDYSPSAGYSPTLPGYSPSSTGQYTPHEGEKNAKTGK-DVSKDDKSNP----      1839
tr|V4P2F5|V4P2F5_EUTSA              SPDYSPSAGYSPTLPGYSPSSTGQYTPHEGDKNDKSGK-DGNKDDKSNP----      1837
tr|R0GXL7|R0GXL7_9BRAS              SPDYSPSAGYSPTLPGYSPSSTGQYTPHEGDKNDKTGK-DVSKNDKSNP----      1838
tr|A0A6J0MG42|A0A6J0MG42_RAPSA      SPDYSPSAGYSPTLPGYSTSSTGQYTPHEGDKNDKTGKD-ASKYGKSNP----      1848
tr|A0A398AHR5|A0A398AHR5_BRACM      SPDYSPNAGYSPTLPGYSSSSTGQYTPHEGYENDKTGKD-ASKDGKSNP----      1847
tr|A0A0D3A1P5|A0A0D3A1P5_BRAOL      SPDYSPSAGYSPTLPGYSSSSTGQYTPHEGYENDKTGED-ASKDGKSSP----      1839
tr|A0A3N6QGC8|A0A3N6QGC8_BRACR      ----------------------------------------------------      1587
tr|A0A0E0LRD6|A0A0E0LRD6_ORYPU      IYRY------------------------------------------------      1807
tr|K3YFU7|K3YFU7_SETIT              SPNYSPTGSYSPTAPGYSPSSTGQATNDKDDESA---R---------------      1852
tr|J3MQH6|J3MQH6_ORYBR              SPNYSPSGSYSPTAPGYSPSSTGQQFSPRAANKD---EDDAQ-----------      1855
tr|A0A803MAV7|A0A803MAV7_CHEQI      SPQYSPSAGYSPSAPGYSPSSTSQYTPQMSDKDDKSKR---------------      1843
tr|A0A0K9RSK8|A0A0K9RSK8_SPIOL      SPQYSPSAGYSPSAPGYSPSSTSQYTPQMSDKDDKSKR---------------      1806
tr|A0A6P6TFX2|A0A6P6TFX2_COFAR      SPQYSPSAGYSPTAPGYSPSSTSQYTSRTTERDDKSVKDDRGRR---------      1850
tr|A0A6I9UDD3|A0A6I9UDD3_SESIN      SPQYSPSAGYSPSAPGYSPSSTSQYTPRTNDRDDKSVKDEKSKR---------      1848
tr|A0A1S3XM16|A0A1S3XM16_TOBAC      ----------------------------------------------------      1630
tr|A0A2G3A9T8|A0A2G3A9T8_CAPAN      SPQYSPSAGYSPSAPGYSPSSTSQYTPRISDRDNKSVKDDKTG----------      1846
tr|A0A1U8FXJ4|A0A1U8FXJ4_CAPAN      SPQYSPSAGYSPSAPGYSPSSTSQYTPRISDRDNKSVKDDKTG----------      1846
tr|M0ZVV4|M0ZVV4_SOLTU              SPQYSPSAGYSPSAPGYSPSSTSQYTPRVSERDNKSVKDDKAG----------      1839
tr|A0A6N2BNI8|A0A6N2BNI8_SOLCI      SPQYSPSAGYSPSAPGYSPSSTSQYTPRVSERDNRSVKDDKAG----------      1846
tr|A0A3Q7F7G2|A0A3Q7F7G2_SOLLC      SPQYSPSAGYSPSAPGYSPSSTSQYTPRVSERDNRSVKDDKAG----------      1846
tr|A0A6V7PIB4|A0A6V7PIB4_ANACO      SPQYSPSASYSPTAPGYSPSSTSQYSPQMSNKDDESTHG--------------      1794
tr|A0A6I9RZH4|A0A6I9RZH4_ELAGV      SPQYSPSAGYSPTAPGYSPSSTSQYTPQMSNKDEESTR---------------      1842
```

```
tr|A0A4Y7I7U1|A0A4Y7I7U1_PAPSO      SPQFSPSAGYSPTAPGYSPSSTSQFTPETSNKDDGSTR---------------     1845
tr|A0A6P9E4T1|A0A6P9E4T1_JUGRE      SPPYSPSAGYSPSQPGYSPSSTSQYTPQMSNKDDRSDKDDRSTR---------     1847
tr|A0A7J7HU38|A0A7J7HU38_CAMSI      SPQYSPSVGYSPSAPGYSPSSTSQYTPQMSNKDNGSKR---------------     1859
tr|F6H0D9|F6H0D9_VITVI              SPQYSPSAGYSPSAPGYSPSSTSQYTPQMSNKDNGSPQ---------------     1852
tr|A0A7J6EME3|A0A7J6EME3_CANSA      SPQYSPSAGYSPSQPGYSPTSTSQYTPQTSDKDDNDDKSTR------------     1806
tr|A0A067G7I3|A0A067G7I3_CITSI      SPQYSPSAGYSPSAPGYSPSSTSQYTPQTNRDDS-TTKDDKNTKGDKSSR---     1815
tr|A0A5J5TPI9|A0A5J5TPI9_GOSBA      SPQYSPSAGYSPSAPGYSPTSTSQYTPSN--KDGRSNKD------DRSKR---     1853
tr|A0A061DGE8|A0A061DGE8_THECC      SPQYSPSAGYSPSAPGYSPSSTSQYTPQTSNKDDRATKDDRSSKDDRSKR---     1861
tr|A0A6P5ZFF2|A0A6P5ZFF2_DURZI      SPQYSPSAGYSPSAPGYSPSSTSQYTPS--NKDDHATKDDRNSKDDRSKH---     1859
tr|B9SXC0|B9SXC0_RICCO              SPQYSPSAGYSPSAPGYSPSSTSQYTQ-TSTKDDRTDKGDRNGRDNKDEKSSR     1855
tr|A0A251M2B5|A0A251M2B5_MANES      SPQYSPSAGYSPSAPGYSPSSTSQYTPQTSAKDDRNNKGDRNGRDNKG-----     1858
tr|A0A314Z8L7|A0A314Z8L7_PRUYE      SPQYSPSAGYSPSQPGYSPSSTSQYTPQTSEKDTKDDRSTR------------     1837
tr|M5VSC7|M5VSC7_PRUPE              SPQYSPSAGYSPSQPGYSPSSTSQYTPQTSEKDTKDDRSTR------------     1844
tr|A0A6J5XYF9|A0A6J5XYF9_PRUAR      SPQYSPSAGYSPSQPGYSPSSTSQYTPQTSEKDTKDDRSTR------------     1844
tr|A0A6J1L0Z7|A0A6J1L0Z7_CUCMA      SPQYSPSAGYSPTAPGYSPSSNSQYTPQTSDKDDRSNR---------------     1848
tr|A0A0A0L655|A0A0A0L655_CUCSA      SPQYSPSAGYSPTAPGYSPSSTSQYTPQTSDKDDRSRKDDRNN----------     1867
```

**Fig. 9. MSA of the elongation subunits B1 of RNAP II from plant sources**

*A0A1Z5RIH1_SORBI, Sorghum bicolor*
*A0A3B6MZU3_WHEAT, Triticum aestivum*
*M4D210_BRARP, Brassica rapa*
*A0A1J3F625_NOCCA, Noccaea caerulescens*
*P18616|NRPB1_ARATH, Arabidopsis thaliana*
*V4P2F5_EUTSA, Eutrema salsugineum*
*A0A6J0MG42_RAPSA, Raphanus sativus*
*A0A0D3A1P5_BRAOL, Brassica oleracea*
*A0A0E0LRD6_ORYPU, Oryza punctata*
*J3MQH6_ORYBR, Oryza brachyantha*
*A0A0K9RSK8_SPIOL, Spinacia oleracea*
*A0A6I9UDD3_SESIN, Sesamum indicum*
*A0A2G3A9T8_CAPAN, Capsicum annuum*
*M0ZVV4_SOLTUSolanum tuberosum*
*A0A3Q7F7G2_SOLLC, Solanum lycopersicum*
*A0A6I9RZH4_ELAGV, Elaeis guineensis*
*A0A6P9E4T1_JUGRE, Juglans regia*
*F6H0D9_VITVI, Vitis vinifera*
*A0A067G7I3_CITSI, Citrus sinensis*
*A0A061DGE8_THECC, Theobroma cacao*
*B9SXC0_RICCO, Ricinus communis*
*A0A314Z8L7_PRUYE, Prunus yedoensis*
*A0A6J5XYF9_PRUAR, Prunus armeniaca*
*A0A0A0L655_CUCSA, Cucumis sativus*
*C5Y387_SORBI, Sorghum bicolor*
*A0A565C923_9BRAS, Arabis nemorensis*
*A0A6J0KQ09_RAPSA, Raphanus sativus*
*A0A1J3JR03_NOCCA, Noccaea caerulescens*
*D7MC36_ARALL, Arabidopsis lyrata*
*R0GXL7_9BRAS, Capsella rubella*
*A0A398AHR5_BRACM, Brassica campestris*
*A0A3N6QGC8_BRACR, Brassica cretica*
*K3YFU7_SETIT, Setaria italic*
*A0A803MAV7_CHEQI, Chenopodium quinoa*
*A0A6P6TFX2_COFAR, Coffea Arabica*
*A0A1S3XM16_TOBAC, Nicotiana tabacum*
*A0A1U8FXJ4_CAPAN, Capsicum annuum*
*A0A6N2BNI8_SOLCI, Solanum chilense*
*A0A6V7PIB4_ANACO, Ananas comosus*
*A0A4Y7I7U1_PAPSO, Papaver somniferum*
*A0A7J7HU38_CAMSI, Camellia sinensis*
*A0A7J6EME3_CANSA, Cannabis sativa*
*A0A5J5TPI9_GOSBA, Gossypium barbadense*
*A0A6P5ZFF2_DURZI, Durio zibethinus*
*A0A251M2B5_MANES, Manihot esculenta*
*M5VSC7_PRUPE, Prunus persica*
*A0A6J1L0Z7_CUCMA, Cucurbita maxima*

### 4.3 PR Function in the MSU RNAP III

The RNAP III is the largest and most complex of these multisubunit enzymes and involves in the transcription of genes encoding small, nontranslated RNAs such as tRNAs, 5S rRNA, and U6 snRNA. It is composed of 17 subunits and with a molecular mass of ~ 700 kDa (Table 1). Out of the 17 subunits, 10 are unique to RNAP III and the others are common to two or all three of the RNAPs. The yeast RNAP III binds 1 $Mg^{2+}$ and 5 $Zn^{2+}$ atoms [26]. The second-largest subunit of the RNAP III (C1) is responsible for the transcription of tRNA, 5S RNA and other low molecular weight RNA genes (Table 1). Although the promoters for RNAP I and RNAP II lie upstream of the transcription start site (as in prokaryotes), RNAP III uses 3 types of promoters where two are internal to the gene (Type 1 for 5S rRNA and Type 2 for tRNAs promoters lie downstream of the transcription start site) and one is external to the gene (Type 3 promoters for U6 SnRNA).

Figure 11 shows the MSA analysis of the elongation subunits C1 of RNAP III from different types of yeasts. The template-binding and catalytic pairs are highlighted in yellow and the zinc-binding conserved Cs is highlighted in orange. The completely conserved $Mg^{2+}$-binding site is highlighted in light green. The catalytic pair is slightly different from others as having an invariant M as –R/KM-. The template-binding pair –FG- and the $Mg^{2+}$-binding site with the 3 invariant Ds is conserved in all.
In contrast to RNAP I and RNAP II, three to four S residues are seen immediately followed by the catalytic pair in all and a big 9 amino acid gap is seen between the catalytic amino acid and the first C of the $Zn^{2+}$-binding site.

The PR exonuclease reaction of RNAP III from *S. cerevisiae* was studied by Whitehall et al. [27].

They found that the ribonuclease activity was totally dependent on the presence of a divalent cation and was stimulated by the addition of non-cognate ribonucleotides. $Zn^{2+}$ or $Co^{2+}$ at 7 mM yielded transcripts predominantly shortened by two nucleotides, but not significantly more. Remarkably, the deletion mutant RNAP III Δ, deprived of C11 subunit, lacked the intrinsic RNA cleavage activity of complete RNAP III. Therefore, Chédin et al. [28] have concluded that the C11 subunit is essential for the PR functions and proposed that C11 (= to RpB9 of RNAP II and A12.2 of RNAP I) allows the enzyme to switch between RNA elongation and RNA cleavage modes and plays an essential role of the RNAP III RNA intrinsic cleavage activity.

CLUSTAL O (1.2.4) MSA of the elongation subunits B1 of RNAP II from animal and insect sources

```
tr|W2TT11|W2TT11_NECAM              GLMDERQGVIDRRGRCMTCAGNLADCPGHFAHLELARPVFHIGFLTKILKVLRCVCFYCSKLLL   114
tr|A0A6P8Y2M5|A0A6P8Y2M5_DROAB      GLMDERQGVLDRNTRCQTCSGNMTECPGHFGHIELTKPVFHIGFIAKTIKILRCVCFYCSKLLV   115
tr|A0A1X7VVZ9|A0A1X7VVZ9_AMPQE      GLMDERQGVIDRASRCQTCAGNMNTCPGHFGHISLTKPVFHVCFMTKIVKIMRCVCFYCSRLLI   116
tr|V5HBP6|V5HBP6_IXORI              GLMDERQGVIDRTSRCQTCAGSMTECPGHFGHIELAKPVFHCGFLTKTIKILRCVCFYCSKLLV   119
tr|A0A7J5XJ09|A0A7J5XJ09_DISMA      GLMDERQGVIDRSGRCQTCAGNMTECPGHFGHIELAKPVFHVGFVTKIMKILRCVCFFCSKLLV   117
tr|A0A671PGW1|A0A671PGW1_9TELE      GLMDERQGVIERSGRCQTCAGNMTECPGHFGHIELAKPVFHVGFITKIMKVLRCVCFFCSKLLV   117
tr|A0A3P9D2C8|A0A3P9D2C8_9CICH      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFISKIMKVLRCVCFFCSKLLV   117
tr|A0A1S3SD38|A0A1S3SD38_SALSA      GLMDERQGVIERSGRCQTCAGNMTECPGHFGHIELAKPVFHVGFVNKIMKVLRCVCYSCSKLLV   117
tr|A0A060XGJ1|A0A060XGJ1_ONCMY      GLMDERQGVIERSGRCQTCAGNMTECPGHFGHIELAKPVFHVGFVNKIMKVLRCVCYSCSKLLV   117
tr|A0A1S3N6M0|A0A1S3N6M0_SALSA      GLMDERQGVIERSGRCQTCAGNMTECPGHFGHIELAKPVFHVGFVNKIMKVLRCVCYSCSKLLV   117
tr|A0A3B1J9R7|A0A3B1J9R7_ASTMX      GLMDERQGVIERSGRCQTCAGNMTECPGHFGHIELAKPVFHVGFVTKIMKVLRCVCFFCSKLLV   117
tr|A0A4W4GG80|A0A4W4GG80_ELEEL      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLTKIMKVLRCVCFFCSKLLV   117
tr|A0A484CCJ1|A0A484CCJ1_PERFV      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFVTKIMKVLRCVCFFCSKLLV   117
tr|A0A1A8DQ60|A0A1A8DQ60_9TELE      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFISKIMKILRCVCFFCSKLLV   117
tr|A0A1A8ER05|A0A1A8ER05_9TELE      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFISKIMKILRCVCFFCSKLLV   117
tr|A0A3P9D2C4|A0A3P9D2C4_9CICH      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFISKIMKVLRCVCFFCSKLLV   117
tr|A0A3B5KXA5|A0A3B5KXA5_9TELE      GLMDERQGVIERSGRCQTCAGNMTECPGHFGHIELAKPVFHVGFISKIMKILRCVCFFCSKLLV   117
tr|H2LPT8|H2LPT8_ORYLA              G---ERQGVIERSGRCQTCAGNMTECPGHFGHIELAKPVFHVGFISKIMKVLRCVCFFCSKLLV   117
tr|A0A4W6CUU1|A0A4W6CUU1_LATCA      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFITKIMKVLRCVCFFCSKLLV   117
tr|A0A3L7GMG1|A0A3L7GMG1_CRIGR      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A6P3HSN8|A0A6P3HSN8_BISBI      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A5N3WIL3|A0A5N3WIL3_MUNMU      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A5N4D616|A0A5N4D616_CAMDR      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A452RK12|A0A452RK12_URSAM      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A670ZA05|A0A670ZA05_PSETE      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLGKTMKVLRCVCFFCSKLLV   118
tr|G3RTC9|G3RTC9_GORGO              GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A7M4DUC2|A0A7M4DUC2_PIG        GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A2K6DF42|A0A2K6DF42_MACNE      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A1S2ZNJ1|A0A1S2ZNJ1_ERIEU      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|O08847|O08847_MOUSE              GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A1D5RJ13|A0A1D5RJ13_MACMU      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|G3MZY8|G3MZY8_BOVIN              GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A6P3E8E8|A0A6P3E8E8_SHEEP      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A667HCS4|A0A667HCS4_LYNCA      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
sp|P24928|RPB1_HUMAN                GLMDPRQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|G3TV69|G3TV69_LOXAF              GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|F6Z7Q4|F6Z7Q4_HORSE              GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
sp|P11414|RPB1_CRIGR                GGLMDPRQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A6I9L929|A0A6I9L929_PERMB      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|I3LYD0|I3LYD0_ICTTR              GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A2K6RYV7|A0A2K6RYV7_SAIBB      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|F7IJW5|F7IJW5_CALJA              GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A671F2F4|A0A671F2F4_RHIFE      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A6J0ZAW1|A0A6J0ZAW1_ODOVR      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A6J3I271|A0A6J3I271_SAPAP      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A091CJT9|A0A091CJT9_FUKDA      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|K9J413|K9J413_DESRO              GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A6J2NJJ4|A0A6J2NJJ4_9CHIR      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|G1MCZ1|G1MCZ1_AILME              GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A3Q7XGI2|A0A3Q7XGI2_URSAR      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A2U3Y5R9|A0A2U3Y5R9_LEPWE      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A6J2AXC2|A0A6J2AXC2_ZALCA      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|D4A5A6|D4A5A6_RAT                GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A2I3M9H2|A0A2I3M9H2_PAPAN      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A2I3RTL5|A0A2I3RTL5_PANTR      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A0R4J0V5|A0A0R4J0V5_MOUSE      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A6J1YFT5|A0A6J1YFT5_ACIJB      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|G3T277|G3T277_LOXAF              GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|H0VRW6|H0VRW6_CAVPO              GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A480L5T3|A0A480L5T3_PIG        GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
sp|P08775|RPB1_MOUSE                GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
tr|A0A340X8C7|A0A340X8C7_LIPVE      GLMDERQGVIERTGRCQTCAGNMTECPGHFGHIELAKPVFHVGFLVKTMKVLRCVCFFCSKLLV   119
                                    ********:*  ** **:*.:  *****.*:.*::****  *: * :*::****: **:**:
```

```
//
tr|W2TT11|W2TT11_NECAM          HRVKILPWSTFRMNLSVTTPYNADFDGDEMNLHLPQSLETRAEIEEIAMVPRQLITPQAN  523
tr|A0A6P8Y2M5|A0A6P8Y2M5_DROAB  HRVKVLPWSTFRMNLSCTSPYNADFDGDEMNLHVPQSMETRAEVENIHITPRQIITPQAN  522
tr|A0A1X7VVZ9|A0A1X7VVZ9_AMPQE  HRVKILPWSTFRLNLSVTTPYNADFDGDEMNLHVPQSLETKAEIQELMMVHRNILTPQSN  524
tr|V5HBP6|V5HBP6_IXORI          HRIKVLPWSTFRMNLSVTTPYNADFDGDEMNLHVPQSLETRAEVEQLAMVPRQIITPQSN  530
tr|A0A7J5XJ09|A0A7J5XJ09_DISMA  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  529
tr|A0A671PGW1|A0A671PGW1_9TELE  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  529
tr|A0A3P9D2C8|A0A3P9D2C8_9CICH  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  529
tr|A0A1S3SD38|A0A1S3SD38_SALSA  HRVRILPWSTFRMNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  530
tr|A0A060XGJ1|A0A060XGJ1_ONCMY  HRVRILPWSTFRMNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  527
tr|A0A1S3N6M0|A0A1S3N6M0_SALSA  HRVRILPWSTFRMNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  527
tr|A0A3B1J9R7|A0A3B1J9R7_ASTMX  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  529
tr|A0A4W4GG80|A0A4W4GG80_ELEEL  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  529
tr|A0A484CCJ1|A0A484CCJ1_PERFV  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  529
tr|A0A1A8DQ60|A0A1A8DQ60_9TELE  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  529
tr|A0A1A8ER05|A0A1A8ER05_9TELE  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  529
tr|A0A3P9D2C4|A0A3P9D2C4_9CICH  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  529
tr|A0A3B5KXA5|A0A3B5KXA5_9TELE  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  529
tr|H2LPT8|H2LPT8_ORYLA          HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  529
tr|A0A4W6CUU1|A0A4W6CUU1_LATCA  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  529
tr|A0A3L7GMG1|A0A3L7GMG1_CRIGR  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  495
tr|A0A6P3HSN8|A0A6P3HSN8_BISBI  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A5N3WIL3|A0A5N3WIL3_MUNMU  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A5N4D616|A0A5N4D616_CAMDR  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  537
tr|A0A452RK12|A0A452RK12_URSAM  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A670ZA05|A0A670ZA05_PSETE  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  530
tr|G3RTC9|G3RTC9_GORGO          HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A7M4DUC2|A0A7M4DUC2_PIG    HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A2K6DF42|A0A2K6DF42_MACNE  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A1S2ZNJ1|A0A1S2ZNJ1_ERIEU  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|O08847|O08847_MOUSE          HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A1D5RJ13|A0A1D5RJ13_MACMU  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|G3MZY8|G3MZY8_BOVIN          HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A6P3E8E8|A0A6P3E8E8_SHEEP  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A667HCS4|A0A667HCS4_LYNCA  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
sp|P24928|RPB1_HUMAN            HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|G3TV69|G3TV69_LOXAF          HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|F6Z7Q4|F6Z7Q4_HORSE          HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
sp|P11414|RPB1_CRIGR            HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A6I9L929|A0A6I9L929_PERMB  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|I3LYD0|I3LYD0_ICTTR          HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A2K6RYV7|A0A2K6RYV7_SAIBB  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|F7IJW5|F7IJW5_CALJA          HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A671F2F4|A0A671F2F4_RHIFE  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A6J0ZAW1|A0A6J0ZAW1_ODOVR  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A6J3I271|A0A6J3I271_SAPAP  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A091CJT9|A0A091CJT9_FUKDA  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|K9J413|K9J413_DESRO          HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A6J2NJJ4|A0A6J2NJJ4_9CHIR  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|G1MCZ1|G1MCZ1_AILME          HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A3Q7XGI2|A0A3Q7XGI2_URSAR  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A2U3Y5R9|A0A2U3Y5R9_LEPWE  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A6J2AXC2|A0A6J2AXC2_ZALCA  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|D4A5A6|D4A5A6_RAT            HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A2I3M9H2|A0A2I3M9H2_PAPAN  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A2I3RTL5|A0A2I3RTL5_PANTR  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A0R4J0V5|A0A0R4J0V5_MOUSE  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A6J1YFT5|A0A6J1YFT5_ACIJB  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|G3T277|G3T277_LOXAF          HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|H0VRW6|H0VRW6_CAVPO          HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A480L5T3|A0A480L5T3_PIG    HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
sp|P08775|RPB1_MOUSE            HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
tr|A0A340X8C7|A0A340X8C7_LIPVE  HRVRILPWSTFRLNLSVTTPYNADFDGDEMNLHLPQSLETRAEIQELAMVPRMIVTPQSN  531
                                **:::*******:*** *:*********:*****:***:**:**:::: :. * ::***:*
```

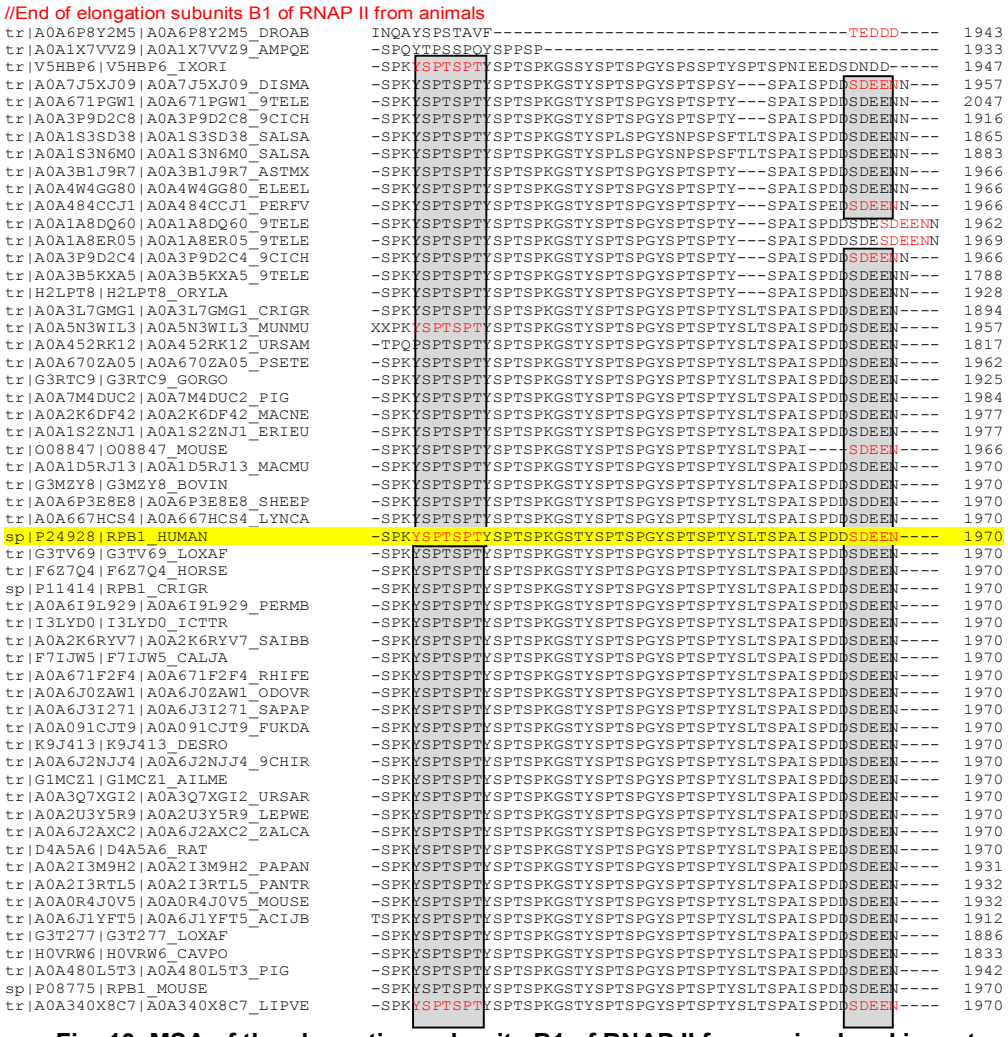



**Fig. 10. MSA of the elongation subunits B1 of RNAP II from animal and insect sources**

*A0A6P8Y2M5_DROAB Drosophila albomicans*
*A0A1X7VVZ9_AMPQE Amphimedon queenslandica*
*V5HBP6_IXORI Ixodes ricinus*
*A0A7J5XJ09_DISMA Dissostichus mawsoni*
*A0A671F2F4_RHIFE Rhinolophus ferrumequinum*
*A0A3P9D2C8_9CICH Maylandia zebra*
*A0A1S3N6M0_SALSA Salmo salar*
*A0A060XGJ1_ONCMY Oncorhynchus mykiss*
*A0A1S3SD38_SALSA Salmo salar*
*A0A3B1J9R7_ASTMX Astyanax mexicanus*
*A0A4W4GG80_ELEEL Electrophorus electricus*
*A0A484CCJ1_PERFV Perca flavescens*
*A0A1A8DQ60_9TELE Nothobranchius kadleci*
*A0A1A8ER05_9TELE Nothobranchius korthausae*
*A0A3P9D2C4_9CICH Maylandia zebra*
*A0A3B5KXA5_9TELE Xiphophorus couchianus*
*H2LPT8_ORYLA Oryzias latipes*
*A0A4W6CUU1_LATCA Lates calcarifer*
*A0A3L7GMG1_CRIGR Cricetulus griseus*
*A0A6P3HSN8_BISBI Bison bison bison*
*A0A5N3WIL3_ Muntiacus muntjak*
*A0A5N4D616_CAMDR Camelus dromedaries*
*A0A452RK12_URSAM Ursus americanus*
*A0A670ZA05_PSETE Pseudonaja textilis*
*G3RTC9_GORGO Gorilla gorilla gorilla*
*H0VRW6_CAVPO Cavia porcellus*
*A0A7M4DUC2_PIG Sus scrofa*
*A0A2K6RYV7_SAIBB Saimiri boliviensis boliviensis*
*F7IJW5_CALJA Callithrix jacchus*
*A0A1S2ZNJ1_ERIEU Erinaceus europaeus*
*O08847_MOUSE Mus musculus*
*A0A1D5RJ13_MACMU Macaca mulatta*
*G3MZY8_BOVIN Bos Taurus*
*A0A6P3E8E8_SHEEP Ovis aries*
*A0A667HCS4_LYNCA Lynx Canadensis*
*P24928|RPB1_HUMAN Homo sapiens*
*G3TV69_LOXAF Loxodonta Africana*
*F6Z7Q4_HORSE Equus caballus*
*P11414|RPB1_CRIGR Cricetulus griseus*
*A0A6I9L929_PERMB Peromyscus maniculatus bairdii*
*I3LYD0_ICTTR Ictidomys tridecemlineatus*
*A0A2K6DF42_MACNE Macaca nemestrina*
*A0A671PGW1_9TELE Sinocyclocheilus anshuiensis*
*A0A6J0ZAW1_ODOVR, Odocoileus vir. texanus*

*A0A6J3I271_SAPAP, Sapajus paella*
*K9J413_DESRO Desmodus rotundus*
*G1MCZ1_AILME Ailuropoda melanoleuca*
*A0A2U3Y5R9_LEPWE Leptonychotes weddellii*
*D4A5A6_RAT Rattus norvegicus*
*A0A2I3RTL5_PANTR Pan troglodytes*
*A0A6J1YFT5_ACIJB Acinonyx jubatus*
*A0A480L5T3_PIG Sus scrofa*
*A0A340X8C7_LIPVE Lipotes vexillifer*
*A0A091CJT9_FUKDA, Fukomys damarensis*
*A0A6J2NJJ4_9CHIR, Phyllostomus discolor*
*A0A3Q7XGI2_URSAR Ursus arctos horribilis*
*A0A6J2AXC2_ZALCA Zalophus californianus*
*A0A2I3M9H2_PAPAN Papio anub*
*A0A0R4J0V5_MOUSE Mus musculus*
*G3T277_LOXAF Loxodonta Africana*
*P08775|RPB1_MOUSE Mus musculus*

CLUSTAL O (1.2.4) MSA of elongation subunits C1 of RNAP III from different yeasts

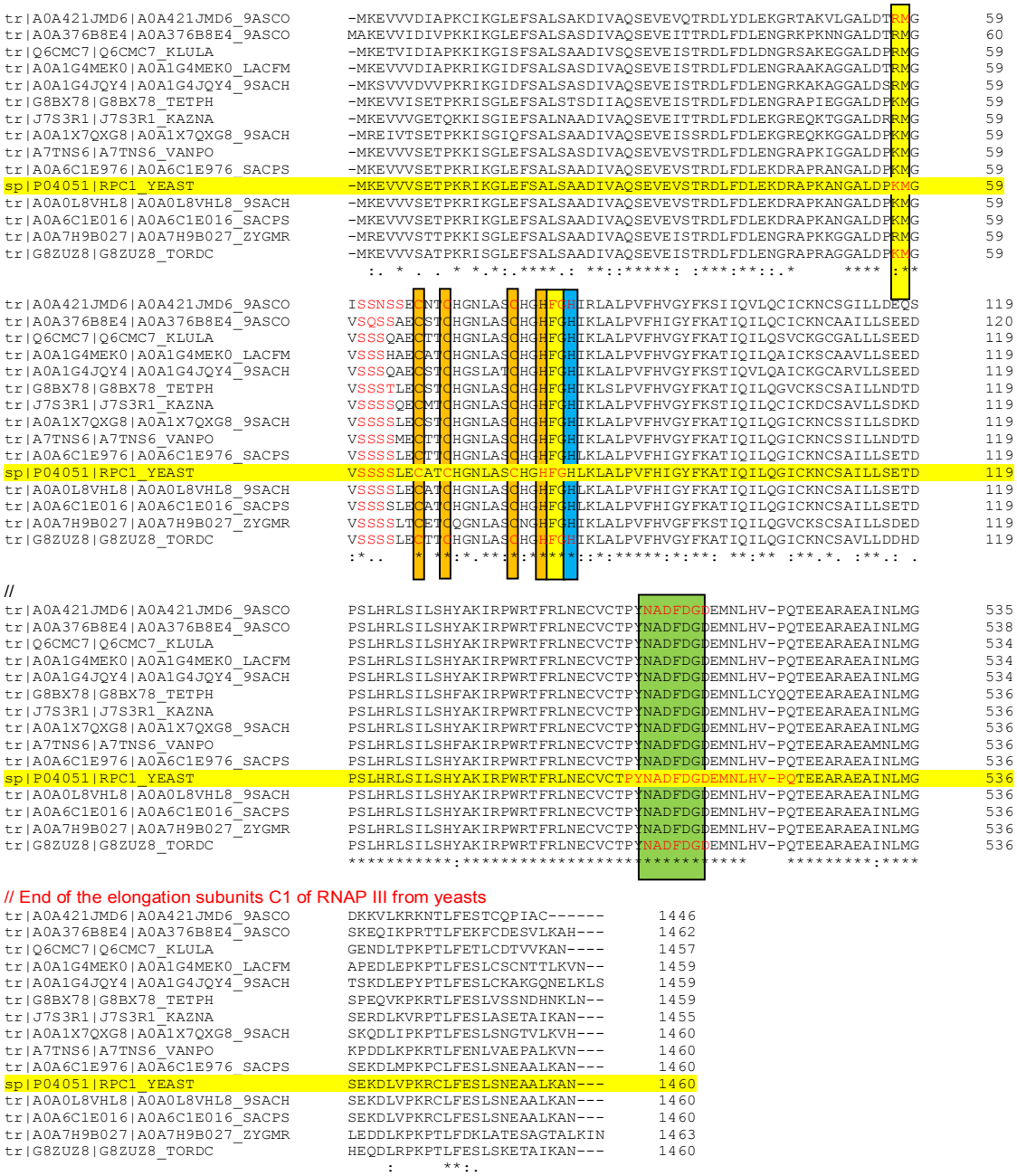


**Fig. 11. MSA of elongation subunits C1of RNAP III from different yeasts**

*A0A421JMD6_9ASCO Spathaspora* sp.
*Q6CMC7_KLULA Kluyveromyces lactis*
*A0A1G4JQY4_9SACH Lachancea mirantina*
*A0A1X7QXG8_9SACH Kazachstania saulgeensis*
*A7TNS6_VANPO Vanderwaltozyma polyspora*
*P04051|RPC1_YEAST Saccharomyces cerevisiae*
*A0A6C1E016_SACPS Saccharomyces pastorianus*
*G8ZUZ8_TORDC Torulaspora delbrueckii*
*A0A376B8E4_9ASCO Saccharomycodes ludwigii*
*A0A1G4MEK0_LACFM Lachancea fermentati*
*J7S3R1_KAZNA Kazachstania naganishii*
*G8BX78_TETPH Tetrapisispora phaffii*
*A0A6C1E976_SACPS Saccharomyces pastorianus*
*A0A0L8VHL8_9SACH Saccharomyces boulardii*
*A0A7H9B027_ZYGMR Zygotorulaspora mrakii*

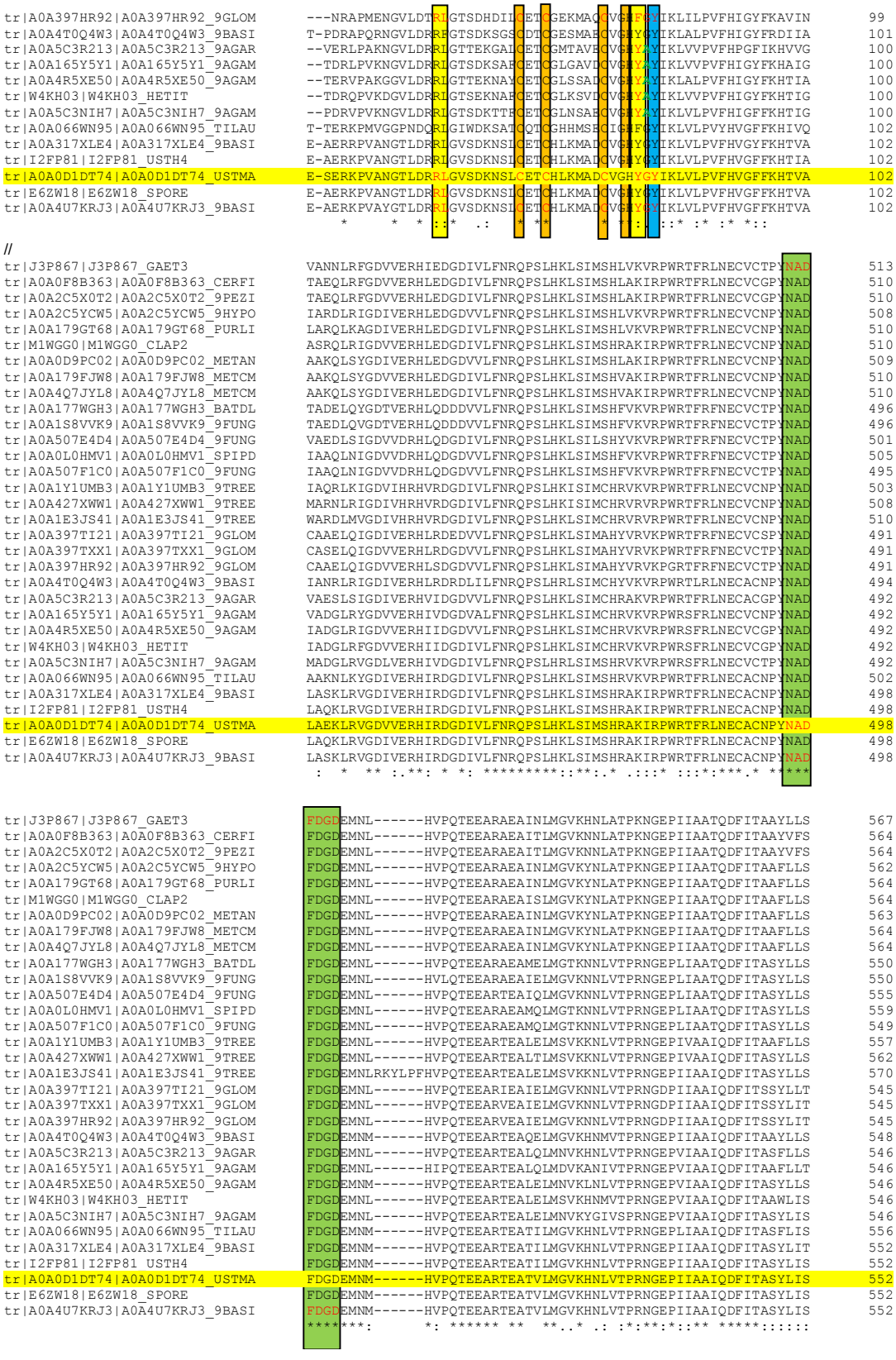
Figure 12 shows the MSA analysis of the elongation subunits C1 of RNAP III from higher fungi. The template-binding and catalytic pairs are highlighted in yellow and the Zn-binding conserved Cs is highlighted in orange. The completely conserved $Mg^{2+}$-binding site is highlighted in light green. The catalytic amino acid –K/H- is conserved in all whereas the template-binding pair is varied as –FG/YG/YA- and the $Mg^{2+}$-binding site with the 3 invariant Ds are conserved in all. A 9 amino acid-gap is seen between the catalytic amino acid and the first C of the $Zn^{2+}$-binding site as in yeasts. The Mg2+-binding invariant Ds are completely conserved in all (Fig. 12). The invasive fungal pathogens (highlighted in light blue) use either a V or an I (highlighted in red) immediately after the template-binding pair, –FG-. *Powellomyces hirtus, Batrachochytrium salamandrivorans, Batrachochytrium dendrobatidis, Spizellomyces punctatus, Spizellomyces palustris* are zoosporic (i.e., reproduce with zoospores) invasive fungal pathogens and belong to Chytridiomycota.

Figure 13 shows the MSA analysis of the elongation subunits C1 of RNAP III from various plant sources. The template-binding and catalytic pairs are highlighted in yellow and the Zn-binding conserved Cs is highlighted in orange. The completely conserved $Mg^{2+}$ binding site is highlighted in light green. The catalytic amino acid –K- is conserved in all, whereas the template-binding pair is varied as –Y/FG- and the $Mg^{2+}$-binding site with the 3 invariant Ds are conserved in all. The distance conservation between the catalytic amino acid and the first C of the $Zn^{2+}$-binding site is maintained in plant sources also as in yeasts and higher fungi.

Figure 14 shows the MSA analysis of the elongation subunits C1 of RNAP III from animals and animal parasites. The template-binding and catalytic pairs are highlighted in yellow and the Zn-binding conserved Cs is highlighted in orange. The completely conserved $Mg^{2+}$-binding site is highlighted in light green. The catalytic amino acid -R- is conserved in all except in the malarial parasite, where an equivalent amino acid -K- is used. Similarly, the template-binding pair is found to be different in the sequences from the animal pathogen/parasites where they mainly use –FG/WG- instead of an -YG- used by others (the parasite sequence are highlighted in light blue). It is not clear whether these changes in the template-binding pairs have any advantage for the parasites to bind and transcribe differentially from their host machinery. However, the $Mg^{2+}$-binding site with the 3 invariant Ds is conserved in all.

CLUSTAL O (1.2.4) MSA of elongation subunits C1 of RNAP III from higher fungi
(Only the polymerase/PR active site and the $Mg^{2+}$-binding regions are shown)

```
tr|J3P867|J3P867_GAET3          N-N-RTPSRHGPLDHRLGTSSKTAECQTLAQPLQNLSGHFGHVRLPLPAFHIGYLRFVQT      112
tr|A0A0F8B363|A0A0F8B363_CERFI  K-N-RTPYLHGPLDERLGTSSKMSKCSTCNELLQTCMGHFGRVKLPLPCFHIGYLGFIIT      109
tr|A0A2C5X0T2|A0A2C5X0T2_9PEZI  K-N-RTPYLHGPLDERLGTSSKMSKCSTCNELLQTCMGHFGRVKLPLPCFHIGYLGFIIT      109
tr|A0A2C5YCW5|A0A2C5YCW5_9HYPO  K-N-RAPYRHGPLDERLGTSSKIGRCATCQDALKDCIGHFGHVRLPLPAFHIGYLRFVMS      107
tr|A0A179GT68|A0A179GT68_PURLI  N-N-RAPYRHGPLDERLGTSSKIGKCATCQDSLQNCTGHFGHVRLPLPAFHIGYLRFVMS      109
tr|M1WGG0|M1WGG0_CLAP2          N-N-RSPYRHGPLDERLGTSSKIGKCETCQDSLTNCIGHFGHVRLPLPAFHIGYLRFIMS      109
tr|A0A0D9PC02|A0A0D9PC02_METAN  N-N-RAPYRHGPLDERLGTSSKSGKCDTCHDSLQNCIGHFGHVRLPLPAFHIGYLRFIMS      109
tr|A0A179FJW8|A0A179FJW8_METCM  N-N-RAPYRHGPLDERLGTSSKGGKCDTCQDSLQNCIGHFGHVRLPLPAFHIGYLRFIMS      109
tr|A0A4Q7JYL8|A0A4Q7JYL8_METCM  N-N-RAPYRHGPLDERLGTSSKGGKCDTCQDSLQNCIGHFGHVRLPLPAFHIGYLRFIMS      109
tr|A0A177WGH3|A0A177WGH3_BATDL  K-NNRPPAKFGALDERLGTADKSIRCETCGETMQDCVGHFGTIRLVLPVFHIGYFKLMIT      102
tr|A0A1S8VVK9|A0A1S8VVK9_9FUNG  K-TTRPPAKFGALDERLGTADKSIRCDTCGENIQECVGHFGTIRLALPVFHIGYFKLMIT      102
tr|A0A507E4D4|A0A507E4D4_9FUNG  Q-SHRPSVKFGVLDDRLGVADKVKKCETCGEGLQDCIGHFGVVRLAVPVFHIGYFKLMIQ      108
tr|A0A0L0HMV1|A0A0L0HMV1_SPIPD  Q-PNRPSVKFGVLDDRLGTADKMGKCETCGLGLQDCVGHFGVVRLVVPVFHIGYFKLMLS      112
tr|A0A507F1C0|A0A507F1C0_9FUNG  Q-PNRPSVKFGVLDDRLGTADKMGKCETCGLGLQDCVGHFGVVRLVVPVFHIGYFKLMLS      102
tr|A0A1Y1UMB3|A0A1Y1UMB3_9TREE  EDGTRTTAKDGPLDARMGPNDKTTHCHTCGEEATKCVGHYGYIKLALPVFHIGYFRPTIN      110
tr|A0A427XWW1|A0A427XWW1_9TREE  EDGSRTTAPHGPLDARLGPNEKGGICTTCHEDHTHCVGHYGFVKLALPVFHIGYFRPTIN      113
tr|A0A1E3JS41|A0A1E3JS41_9TREE  EDGSRTVASHGPIDARMGPNEKGKKCATCGEENAKCVGHYGYIKLVLPVFHIGYFRATIN      113
tr|A0A397TI21|A0A397TI21_9GLOM  ---NRKPMENGVLDTRLGTSDHDTLCQTCGEKMAQCIGHFGYVKLILPVFHIGYFKAVIN      99
tr|A0A397TXX1|A0A397TXX1_9GLOM  ---NRAPMENGVLDTRLGTSDHDILCETCGERMAQCIGHFGYVKLVLPVFHIGYFKAVIN      99
```



```
// End of the elongation subunits C1 of RNAP III from higher fungi
tr|J3P867|J3P867_GAET3                -----------     1461
tr|A0A0F8B363|A0A0F8B363_CERFI        HSHGGDVVMVG     1494
tr|A0A2C5X0T2|A0A2C5X0T2_9PEZI        HSHGGDVVMVG     1493
tr|A0A2C5YCW5|A0A2C5YCW5_9HYPO        -----------     1452
tr|A0A179GT68|A0A179GT68_PURLI        -----------     1446
tr|M1WGG0|M1WGG0_CLAP2                -----------     1450
tr|A0A0D9PC02|A0A0D9PC02_METAN        -----------     1447
tr|A0A179FJW8|A0A179FJW8_METCM        -----------     1448
tr|A0A4Q7JYL8|A0A4Q7JYL8_METCM        -----------     1448
tr|A0A177WGH3|A0A177WGH3_BATDL        -----------     1336
tr|A0A1S8VVK9|A0A1S8VVK9_9FUNG        -----------     1409
tr|A0A507E4D4|A0A507E4D4_9FUNG        -----------     1442
tr|A0A0L0HMV1|A0A0L0HMV1_SPIPD        -----------     1434
tr|A0A507F1C0|A0A507F1C0_9FUNG        -----------     1424
tr|A0A1Y1UMB3|A0A1Y1UMB3_9TREE        -----------     1423
tr|A0A427XWW1|A0A427XWW1_9TREE        -----------     1467
tr|A0A1E3JS41|A0A1E3JS41_9TREE        -----------     1442
tr|A0A397TI21|A0A397TI21_9GLOM        -----------     1415
tr|A0A397TXX1|A0A397TXX1_9GLOM        -----------     1413
tr|A0A397HR92|A0A397HR92_9GLOM        -----------     1417
tr|A0A4T0Q4W3|A0A4T0Q4W3_9BASI        -----------     1425
tr|A0A5C3R213|A0A5C3R213_9AGAR        -----------     1404
tr|A0A165Y5Y1|A0A165Y5Y1_9AGAM        -----------     1404
tr|A0A4R5XE50|A0A4R5XE50_9AGAM        -----------     1404
tr|W4KH03|W4KH03_HETIT                -----------     1403
tr|A0A5C3NIH7|A0A5C3NIH7_9AGAM        -----------     1404
tr|A0A066WN95|A0A066WN95_TILAU        -----------     1404
tr|A0A317XLE4|A0A317XLE4_9BASI        -----------     1400
tr|I2FP81|I2FP81_USTH4                -----------     1400
tr|A0A0D1DT74|A0A0D1DT74_USTMA        -----------     1400
tr|E6ZW18|E6ZW18_SPORE                -----------     1400
tr|A0A4U7KRJ3|A0A4U7KRJ3_9BASI        -----------     1400
```

**Fig. 12. MSA of elongation subunits C1 of RNAP III from higher fungi**

*J3P867_GAET3 Gaeumannomyces tritici*
*A0A0F8B363_CERFI Ceratocystis fimbriata*
*A0A2C5YCW5_9HYPO Cordyceps sp.*
*A0A2C5X0T2_9PEZI Ceratocystis fimbriata*
*A0A179GT68_PURLI Purpureocillium lilacinum*
*M1WGG0_CLAP2 Claviceps purpurea*
*A0A0D9PC02_METAN Metarhizium anisopliae*
*A0A179FJW8_METCM Pochonia chlamydosporia*
*A0A4Q7JYL8_METCM Pochonia chlamydosporia*
*A0A177WGH3_BATDL Batrachochytrium dendrobatidis*
*A0A1S8VVK9_9FUNG B. salamandrivorans*
*A0A507E4D4_9FUNG Powellomyces hirtus*
*A0A0L0HMV1_SPIPD Spizellomyces punctatus*
*A0A507F1C0_9FUNG Spizellomyces sp. 'palustris'*
*A0A1Y1UMB3_9TREE Kockovaella imperatae*
*A0A427XWW1_9TREE Apiotrichum porosum*
*A0A1E3JS41_9TREE Cryptococcus depauperatus*
*A0A397TI21_9GLOM Glomus cerebriforme*
*A0A397TXX1_9GLOM Gigaspora rosea*
*A0A397HR92_9GLOM Diversispora epigaea*
*A0A4T0Q4W3_9BASI Wallemia mellicola*
*A0A5C3R213_9AGAR Pterula gracilis*
*A0A165Y5Y1_9AGAM Peniophora sp.*
*A0A4R5XE50_9AGAM Rickenella mellea*
*W4KH03_HETIT Heterobasidion irregular*
*A0A5C3NIH7_9AGAM Heliocybe sulcata*
*A0A066WN95_TILAU Tilletiaria anomala*
*A0A317XLE4_9BASI Testicularia cyperi*
*I2FP81_USTH4 Ustilago hordei*
*A0A0D1DT74_USTMA Ustilago maydis*
*E6ZW18_SPORE Sporisorium reilianum*
*A0A4U7KRJ3_9BASI Sporisorium graminicola*

CLUSTAL O (1.2.4) MSA of elongation subunits C1 of RNAP III from various plant sources (Only the polymerase/PR active site and the $Mg^{2+}$-binding regions are shown)

```
tr|A0A6V1TNL5|A0A6V1TNL5_THAWE      RDLFTMPQRSPAINGVLDPRLGVSDKV-STCSTCKLKLADCAGHFGYIRLALPCFHIGYI       115
tr|A0A6V2I8W3|A0A6V2I8W3_TETST      RGLYKMPERKPLPNGILDTRLGVSSKQ-LACGTCSNKLADCAGHFGYLKLQMPVFHIGYF       116
tr|A0A2R6WX06|A0A2R6WX06_MARPO      SSLYRMPERQPAAHGPLDTRLGTTNKK-GVCTTCHGKLAECPGHFGYVRLELPVFHIGYI       111
sp|F4JXF9|NRPC1_ARATH               IGLYD-HSFKPYENGLLDPRMGPPNKK-SICTTCEGNFQNCPGHYGYLKLDLPVYNVGYF       104
tr|A0A6J0MCR5|A0A6J0MCR5_RAPSA      NMYYE-SNFKPIEGGLLDPRMGPPNKR-STCATCHGNFQNCPGHYGYLKLDLPVYNVGFL       104
tr|A0A3P6CYB3|A0A3P6CYB3_BRAOL      NMYYE-SNFKPIEGGLLDPRMGPPNKR-STCATCHGNFQNCPGHYGYLKLDLPVYNVGFF       104
tr|A0A3N6S603|A0A3N6S603_BRACR      NMYYE-SNFKPIEGGLLDPRMGPPNKR-STCATCHGNFQNCPGHYGYLKLDLPVYNVGFF       104
tr|A0A078F415|A0A078F415_BRANA      NMYYE-SNFKPIEGGLLDPRMGPPNKR-STCATCHGNFQNCPGHYGYLKLDLPVYNVGFF       104
tr|A0A218W7K5|A0A218W7K5_PUNGR      GVYYA-PNRRPIEGGLLDSRMGPPNKG-SRCSTCDADFNGCPGHYGYLNLVLPVYNVGFL       104
tr|V7AFW1|V7AFW1_PHAVU              GSYYD-SFKKPIHGGLLDPRMGPANKS-LGCATCHGNFHDCPGHYGYLNLALPVFNVGYL       106
tr|Q7FAC6|Q7FAC6_ORYSJ              SRIYN-HEMKPVPNGLLDTRMGAANKL-GECSTCHGSFAECPGHFGYLKLALPVFNVGFF       105
tr|A0A1U8AS29|A0A1U8AS29_NELNU      SVYYD-VNRKPIDDGLLDSRMGPASKS-GICTTCDGSFTDCPGHFGYLNLALPVFNVGYF       105
tr|B9T0Q9|B9T0Q9_RICCO              GSYYD-STRKPIQAGLLDPRLGPATKQSSACETCGADFHECPGHFGYLNLVLPVFNVGYM       76
tr|A0A7J8R6T0|A0A7J8R6T0_GOSDV      GVYYD-LQSRPIEGGLLDPRMGPPNKS-GKCATCDGSFGDCPGHYGYLPLVLPVYNVGYL       105
tr|A0A061FU60|A0A061FU60_THECC      ALYYD-PKSRPIEGGLLDPRMGPANKS-GKCATCHGNFADCPGHYGYLSLALPVYNVGYL       100
tr|A0A1R3HB58|A0A1R3HB58_COCAP      DAYYD-ANNRAIVGGLLDPRMGPANKSSPDCATCGAKFVDCPGHYGYLPLALPVYNVGYL       106
                                       :          * ** *:*   *     * **  .:  * **:**: * :* :::*::
```

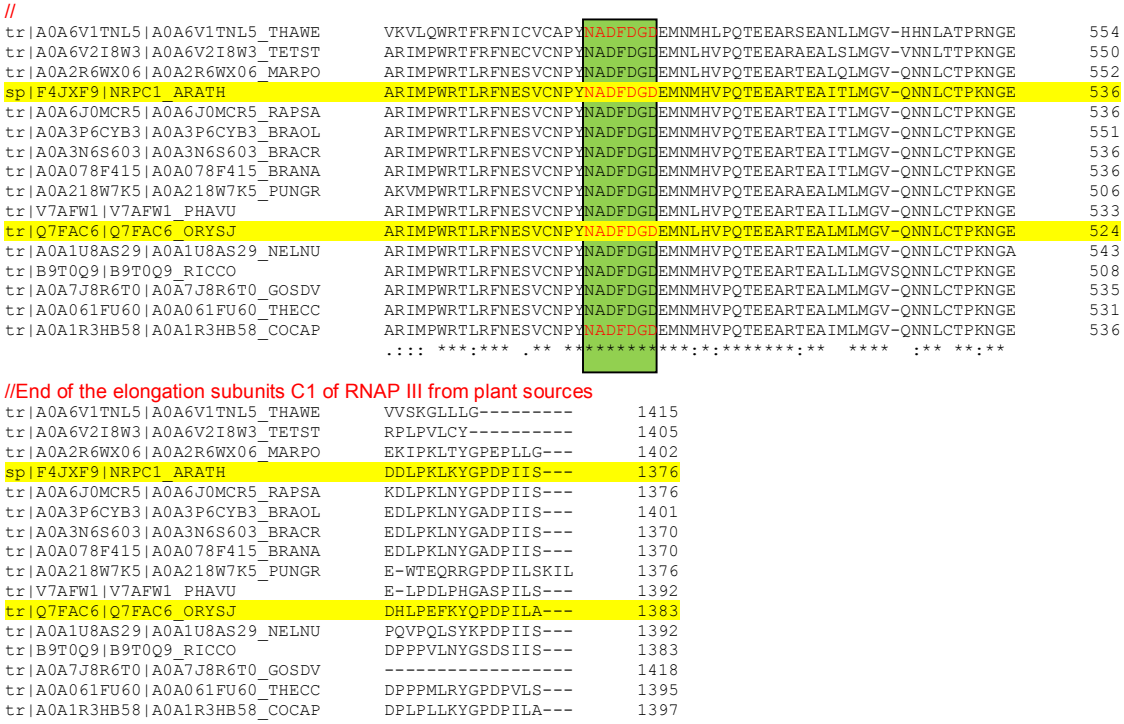
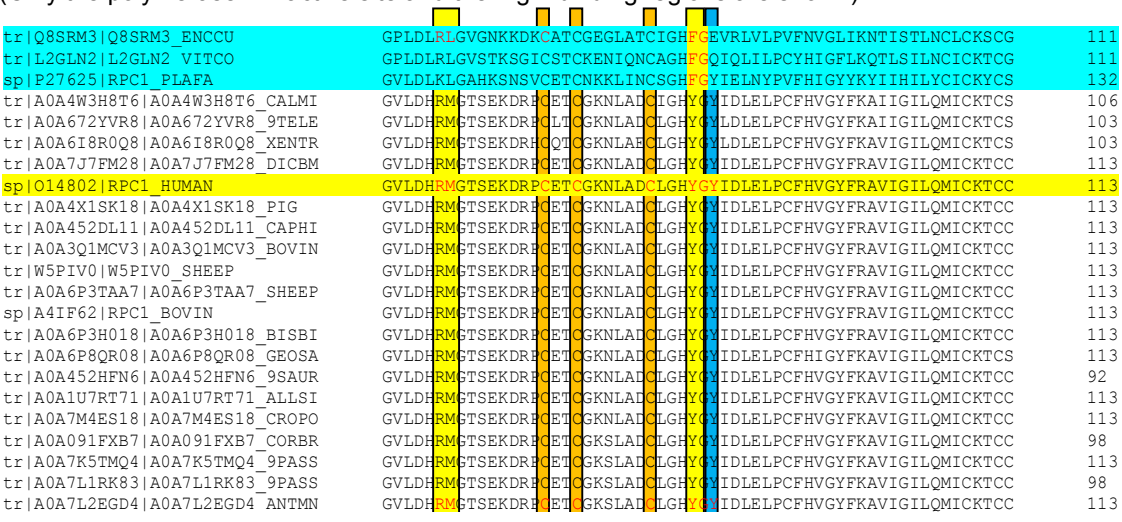
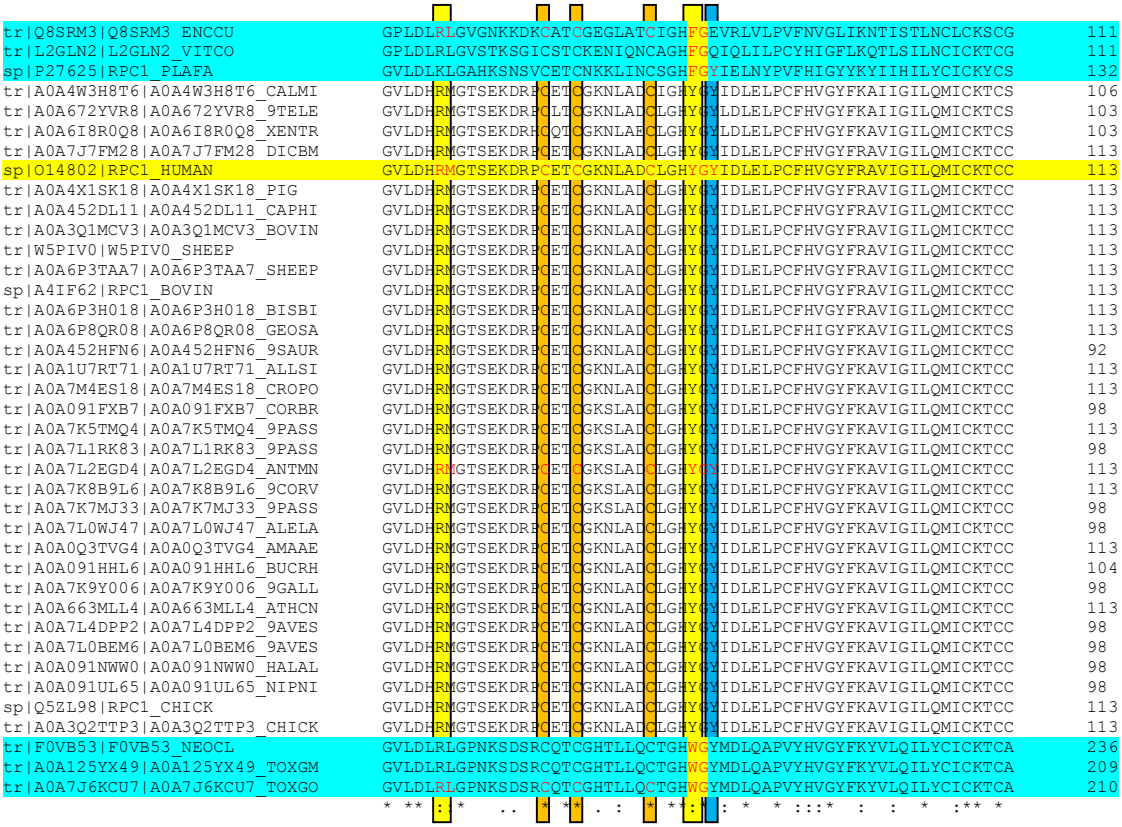



**Fig. 13. MSA of elongation subunits C1 of RNAP III from various plant sources**

*A0A6V1TNL5_THAWE Thalassiosira weissflogii*
*A0A6V2I8W3_TETST Tetraselmis striata*
*A0A2R6WX06_MARPO Marchantia polymorpha*
*F4JXF9|NRPC1_ARATH Arabidopsis thaliana*
*A0A6J0MCR5_RAPSA Raphanus sativus*
*A0A3P6CYB3_BRAOL Brassica oleracea*
*A0A3N6S603_BRACR Brassica cretica*
*A0A078F415_BRANA Brassica napus*
*A0A218W7K5_PUNGR Punica granatum*
*V7AFW1_PHAVU Phaseolus vulgaris*
*Q7FAC6_ORYSJ Oryza sativa subsp. Japonica*
*A0A1U8AS29_NELNU Nelumbo nucifera*
*B9T0Q9_RICCO Ricinus communis*
*A0A7J8R6T0_GOSDV Gossypium davidsonii*
*A0A061FU60_THECC Theobroma cacao*
*A0A1R3HB58_COCAP Corchorus capsularis*

CLUSTAL O (1.2.4) MSA of elongation subunits C1 of RNAP III from animals and animal parasites (Only the polymerase/PR active site and the $Mg^{2+}$-binding regions are shown)




tr|Q8SRM3|Q8SRM3_ENCCU GPLDLRLGVGNKKDKCATCGEGLATCIGHFGEVRLVLPVFNVGLIKNTISTLNCLCKSCG 111
tr|L2GLN2|L2GLN2_VITCO GPLDLRLGVSTKSGICSTCKENIQNCAGHFGQIQLILPCYHIGFLKQTLSILNCICKTCG 111
sp|P27625|RPC1_PLAFA GVLDLKLGAHKSNSVCETCNKKLINCSGHFGYIELNYPVFHIGYYKYIIHILYCICKYCS 132
tr|A0A4W3H8T6|A0A4W3H8T6_CALMI GVLDHRMGTSEKDRPCETCGKNLADCIGHYGYIDLELPCFHVGYFKAIIGILQMICKTCS 106
tr|A0A672YVR8|A0A672YVR8_9TELE GVLDHRMGTSEKDRPCLTCGKNLADCLGHYGYLDLELPCFHVGYFKAIIGILQMICKTCS 103
tr|A0A6I8R0Q8|A0A6I8R0Q8_XENTR GVLDHRMGTSEKDRPCQTCGKNLAECLGHYGYLDLELPCFHVGYFKAVIGILQMICKTCS 103
tr|A0A7J7FM28|A0A7J7FM28_DICBM GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFRAVIGILQMICKTCC 113
sp|O14802|RPC1_HUMAN GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFRAVIGILQMICKTCC 113
tr|A0A4X1SK18|A0A4X1SK18_PIG GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFRAVIGILQMICKTCC 113
tr|A0A452DL11|A0A452DL11_CAPHI GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFRAVIGILQMICKTCC 113
tr|A0A3Q1MCV3|A0A3Q1MCV3_BOVIN GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFRAVIGILQMICKTCC 113
tr|W5PIV0|W5PIV0_SHEEP GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFRAVIGILQMICKTCC 113
tr|A0A6P3TAA7|A0A6P3TAA7_SHEEP GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFRAVIGILQMICKTCC 113
sp|A4IF62|RPC1_BOVIN GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFRAVIGILQMICKTCC 113
tr|A0A6P3H018|A0A6P3H018_BISBI GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFRAVIGILQMICKTCC 113
tr|A0A6P8QR08|A0A6P8QR08_GEOSA GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHIGYFKAVIGILQMICKTCS 113
tr|A0A452HFN6|A0A452HFN6_9SAUR GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 92
tr|A0A1U7RT71|A0A1U7RT71_ALLSI GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 113
tr|A0A7M4ES18|A0A7M4ES18_CROPO GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 113
tr|A0A091FXB7|A0A091FXB7_CORBR GVLDHRMGTSEKDRPCETCGKSLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 98
tr|A0A7K5TMQ4|A0A7K5TMQ4_9PASS GVLDHRMGTSEKDRPCETCGKSLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 113
tr|A0A7L1RK83|A0A7L1RK83_9PASS GVLDHRMGTSEKDRPCETCGKSLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 98
tr|A0A7L2EGD4|A0A7L2EGD4_ANTMN GVLDHRMGTSEKDRPCETCGKSLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 113
tr|A0A7K8B9L6|A0A7K8B9L6_9CORV GVLDHRMGTSEKDRPCETCGKSLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 113
tr|A0A7K7MJ33|A0A7K7MJ33_9PASS GVLDHRMGTSEKDRPCETCGKSLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 98
tr|A0A7L0WJ47|A0A7L0WJ47_ALELA GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 98
tr|A0A0Q3TVG4|A0A0Q3TVG4_AMAAE GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 113
tr|A0A091HHL6|A0A091HHL6_BUCRH GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 104
tr|A0A7K9Y006|A0A7K9Y006_9GALL GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 98
tr|A0A663MLL4|A0A663MLL4_ATHCN GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 113
tr|A0A7L4DPP2|A0A7L4DPP2_9AVES GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 98
tr|A0A7L0BEM6|A0A7L0BEM6_9AVES GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 98
tr|A0A091NWW0|A0A091NWW0_HALAL GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 98
tr|A0A091UL65|A0A091UL65_NIPNI GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 98
sp|Q5ZL98|RPC1_CHICK GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 113
tr|A0A3Q2TTP3|A0A3Q2TTP3_CHICK GVLDHRMGTSEKDRPCETCGKNLADCLGHYGYIDLELPCFHVGYFKAVIGILQMICKTCC 113
tr|F0VB53|F0VB53_NEOCL GVLDLRLGPNKSDSRCQTCGHTLLQCTGHWGYMDLQAPVYHVGYFKYVLQILYCICKTCA 236
tr|A0A125YX49|A0A125YX49_TOXGM GVLDLRLGPNKSDSRCQTCGHTLLQCTGHWGYMDLQAPVYHVGYFKYVLQILYCICKTCA 209
tr|A0A7J6KCU7|A0A7J6KCU7_TOXGO GVLDLRLGPNKSDSRCQTCGHTLLQCTGHWGYMDLQAPVYHVGYFKYVLQILYCICKTCA 210
//
tr|Q8SRM3|Q8SRM3_ENCCU NECVCTPYNADFDGDEMNVHVPQTEKARAEASVLMSVSNNIVTPRHGEPIVAATQDFITG 519
tr|L2GLN2|L2GLN2_VITCO NECVCAPYNADFDGDEMNIHFPQTFEARAEAEELMGVKQNICTARNGEPLISCTQDFLTG 509
sp|P27625|RPC1_PLAFA NECVCSPYNADFDGDEMNLDVPQTEEARAEALYLMNVKHNLITPKNGEVIIALTQDFLSA 660
tr|A0A4W3H8T6|A0A4W3H8T6_CALMI NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 541
tr|A0A672YVR8|A0A672YVR8_9TELE NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 538
tr|A0A6I8R0Q8|A0A6I8R0Q8_XENTR NECVCTPYNADFDGDEMNLHLPQTEEAKAEAYVLMGTKANLVTPRNGEPLIAAIQDFLTA 538
tr|A0A7J7FM28|A0A7J7FM28_DICBM NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 548
sp|O14802|RPC1_HUMAN NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 548
tr|A0A4X1SK18|A0A4X1SK18_PIG NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 548
tr|A0A452DL11|A0A452DL11_CAPHI NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 607
tr|A0A3Q1MCV3|A0A3Q1MCV3_BOVIN NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 548
tr|W5PIV0|W5PIV0_SHEEP NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 550
tr|A0A6P3TAA7|A0A6P3TAA7_SHEEP NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 548
sp|A4IF62|RPC1_BOVIN NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 548
tr|A0A6P3H018|A0A6P3H018_BISBI NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 548
tr|A0A6P8QR08|A0A6P8QR08_GEOSA NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 548
tr|A0A452HFN6|A0A452HFN6_9SAUR NECVCTPYNADFDGDEMNLHLPQTEEAKAEALILMGTKANLVTPRNGEPLIAAIQDFLTG 527
tr|A0A1U7RT71|A0A1U7RT71_ALLSI NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 551
tr|A0A7M4ES18|A0A7M4ES18_CROPO NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 548
tr|A0A091FXB7|A0A091FXB7_CORBR NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 533
tr|A0A7K5TMQ4|A0A7K5TMQ4_9PASS NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 555
tr|A0A7L1RK83|A0A7L1RK83_9PASS NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 533
tr|A0A7L2EGD4|A0A7L2EGD4_ANTMN NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 548
tr|A0A7K8B9L6|A0A7K8B9L6_9CORV NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 548
tr|A0A7K7MJ33|A0A7K7MJ33_9PASS NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 533
tr|A0A7L0WJ47|A0A7L0WJ47_ALELA NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 533
tr|A0A0Q3TVG4|A0A0Q3TVG4_AMAAE NECVCTPYNADFDGDEMNLXLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 548
tr|A0A091HHL6|A0A091HHL6_BUCRH NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 539
tr|A0A7K9Y006|A0A7K9Y006_9GALL NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 533
tr|A0A663MLL4|A0A663MLL4_ATHCN NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 548
tr|A0A7L4DPP2|A0A7L4DPP2_9AVES NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 533
tr|A0A7L0BEM6|A0A7L0BEM6_9AVES NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 533
tr|A0A091NWW0|A0A091NWW0_HALAL NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 533
tr|A0A091UL65|A0A091UL65_NIPNI NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 533
sp|Q5ZL98|RPC1_CHICK NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 548
tr|A0A3Q2TTP3|A0A3Q2TTP3_CHICK NECVCTPYNADFDGDEMNLHLPQTEEAKAEALVLMGTKANLVTPRNGEPLIAAIQDFLTG 548
tr|F0VB53|F0VB53_NEOCL NECVCSPYNADFDGDEMNLHLPQTHEARAEALYLMGVVNNLVTPKNGEPLIAATQDFLSS 648
tr|A0A125YX49|A0A125YX49_TOXGM NECVCSPYNADFDGDEMNLHLPQTHEARAEALYLMGVVNNLVTPKNGEPLIAATQDFLSS 621
tr|A0A7J6KCU7|A0A7J6KCU7_TOXGO NECVCSPYNADFDGDEMNLHLPQTHEARAEALYLMGVVNNLVTPKNGEPLIAATQDFLSS 622

```
// End of the elongation subunits C1 of RNAP III from animals and animal parasites
tr|Q8SRM3|Q8SRM3_ENCCU              ------------  1316
tr|L2GLN2|L2GLN2_VITCO              ------------  1442
sp|P27625|RPC1_PLAFA                E--RETAMNY---  2339
tr|A0A4W3H8T6|A0A4W3H8T6_CALMI      --NSEFHIPLVT-  1383
tr|A0A672YVR8|A0A672YVR8_9TELE      --HADFHIPLIT-  1380
tr|A0A6I8R0Q8|A0A6I8R0Q8_XENTR      --HEDFHHPFLNM  1381
tr|A0A7J7FM28|A0A7J7FM28_DICBM      --TNEFHIPLVT-  1380
sp|O14802|RPC1_HUMAN                --TNEFHIPLVT-  1390
tr|A0A4X1SK18|A0A4X1SK18_PIG        --TDEFHIPLVT-  1291
tr|A0A452DL11|A0A452DL11_CAPHI      --TNEFHIPLVT-  1449
tr|A0A3Q1MCV3|A0A3Q1MCV3_BOVIN      --TNEFHIPLVT-  1390
tr|W5PIV0|W5PIV0_SHEEP              --TNEFHIPLVT-  1392
tr|A0A6P3TAA7|A0A6P3TAA7_SHEEP      --TNEFHIPLVT-  1402
sp|A4IF62|RPC1_BOVIN                --TNEFHIPLVT-  1390
tr|A0A6P3H018|A0A6P3H018_BISBI      --TNEFHIPLVT-  1402
tr|A0A6P8QR08|A0A6P8QR08_GEOSA      --NNEFHIPIIT-  1390
tr|A0A452HFN6|A0A452HFN6_9SAUR      ------------  1354
tr|A0A1U7RT71|A0A1U7RT71_ALLSI      --SNEFHIPIIT-  1393
tr|A0A7M4ES18|A0A7M4ES18_CROPO      --SNEFHIPIIT-  1390
tr|A0A091FXB7|A0A091FXB7_CORBR      --HNEFHIPIVT-  1373
tr|A0A7K5TMQ4|A0A7K5TMQ4_9PASS      --HNEFHIPIVT-  1397
tr|A0A7L1RK83|A0A7L1RK83_9PASS      --HNEFHIPIVT-  1375
tr|A0A7L2EGD4|A0A7L2EGD4_ANTMN      --HNEFHIPIVT-  1390
tr|A0A7K8B9L6|A0A7K8B9L6_9CORV      --HNEFHIPIVT-  1390
tr|A0A7K7MJ33|A0A7K7MJ33_9PASS      --HNEFHIPIVT-  1375
tr|A0A7L0WJ47|A0A7L0WJ47_ALELA      --NNEFHIPIVT-  1376
tr|A0A0Q3TVG4|A0A0Q3TVG4_AMAAE      --NSDFHIPIVT-  1390
tr|A0A091HHL6|A0A091HHL6_BUCRH      --NNEFHIPIVT-  1383
tr|A0A7K9Y006|A0A7K9Y006_9GALL      ------------  1327
tr|A0A663MLL4|A0A663MLL4_ATHCN      --NNEFHIPIVT-  1402
tr|A0A7L4DPP2|A0A7L4DPP2_9AVES      --NNEFHIPIVT-  1375
tr|A0A7L0BEM6|A0A7L0BEM6_9AVES      --NNEFHIPIVT-  1375
tr|A0A091NWW0|A0A091NWW0_HALAL      --NNEFHIPIVT-  1375
tr|A0A091UL65|A0A091UL65_NIPNI      --NNEFHIPIVT-  1375
sp|Q5ZL98|RPC1_CHICK                --NNEFHIPIVT-  1390
tr|A0A3Q2TTP3|A0A3Q2TTP3_CHICK      --NNEFHIPIVT-  1402
tr|F0VB53|F0VB53_NEOCL              FGLQEKHAV----  1769
tr|A0A125YX49|A0A125YX49_TOXGM      YGLPEKHTV----  1746
tr|A0A7J6KCU7|A0A7J6KCU7_TOXGO      YGLPEKHTV----  1747
```

**Fig. 14. MSA of C1 elongation subunits of RNAP III from animal and animal parasites**

*Q8SRM3_ENCCU Encephalitozoon cuniculi*
*P27625|RPC1_PLAFA Plasmodium falciparum*
*A0A672YVR8_9TELE Sphaeramia orbicularis*
*A0A6P9CJ55_PANGU Pantherophis guttatus*
*O14802|RPC1_HUMAN Homo sapiens*
*A0A3Q1MCV3_BOVIN Bos Taurus*
*A0A6P3H018_BISBI Bison bison*
*A0A6P3TAA7_SHEEP Ovis aries*
*A0A452HFN6_9SAUR Gopherus agassizii*
*A0A7M4ES18_CROPO Crocodylus porosus*
*A0A7K5TMQ4_9PASS Cephalopterus ornatus*
*A0A7L2EGD4_ANTMN Anthoscopus minutes*
*A0A7K7MJ33_9PASS Brachypodius atriceps*
*A0A0Q3TVG4_AMAAE Amazona aestiva*
*A0A7K9Y006_9GALL Odontophorus gujanensis*
*A0A7L4DPP2_9AVES Eurystomus gularis*
*A0A091NWW0_HALAL Haliaeetus albicilla*
*Q5ZL98|RPC1_CHICK Gallus gallus*
*F0VB53_NEOCL Neospora caninum*
*A0A7J6KCU7_TOXGO Toxoplasma gondii*
*L2GLN2_VITCO Vittaforma corneae*
*A0A4W3H8T6_CALMI Callorhinchus milii*
*A0A6I8R0Q8_XENTR Xenopus tropicalis*
*A0A7J7FM28_DICBM Diceros bicornis minor*
*A0A4X1SK18_PIG Sus scrofa*
*W5PIV0_SHEEP Ovis aries*
*A4IF62|RPC1_BOVIN Bos taurus*
*A0A6P8QR08_GEOSA Geotrypetes seraphini*
*A0A1U7RT71_ALLSI Alligator sinensis*
*A0A091HHL6_BUCRH Buceros rhinoceros silvestris*
*A0A7L1RK83_9PASS Locustella ochotensis*
*A0A7K8B9L6_9CORV Cnemophilus loriae*
*A0A7L0WJ47_ALELA Alectura lathami*
*A0A091HHL6_BUCRH Buceros rhinoceros silvestris*
*A0A663MLL4_ATHCN Athene cunicularia*
*A0A7L0BEM6_9AVES Spizaetus tyrannus*
*A0A091UL65_NIPNI Nipponia Nippon*
*A0A3Q2TTP3_CHICK Gallus gallus*
*A0A125YX49_TOXGM Toxoplasma gondii*

## 4.4 PR Functions in the Plant-Specific RNAPs IV and V

In addition to the three canonical MSU RNAPs, viz. RNAP I, II and III found in all eukaryotes, two non-redundant plant-specific RNA polymerases, viz. RNAP IV and RNAP V have also been reported from many plant species [29]. Plants have evolved these two specialized RNAPs that mainly involve in transcriptional gene silencing (TGS) in plants. Whereas the RNAP IV is required for small interfering RNA (siRNA) biogenesis, the RNAP V transcripts are required for siRNA methylations of RNA-directed DNA methylation (RdDM) loci. Both are localized in the nucleus, composed of 12 subunits each and

are structurally and functionally distinct from other canonical eukaryotic polymerases discussed above. For example, 4 subunits of RNAP IV and 6 subunits of RNAP V are distinct from RNAP II. The largest catalytic subunit 1 and subunit 7 are unique for RNAP II, IV and V and the subunits 2 and 4 are common for IV and V but different from RNAP II. RNAP IV precursor RNAs are only ~30–40 nucleotides (nts) in length, just long enough to encode single 24-nt siRNAs, whereas the RNAP V makes longer transcripts than RNAP IV (i.e.), ~200 nts or longer [30 and references therein]. Furthermore, RNAP IV and V use the same initiation subunit, but differ only in their elongation subunits. However, they are found to be non-essential for viability but required for RNA-mediated TGS and heterochromatin formation in plants [31].

RNAP IV is known to mainly involve in the formation of heterochromatin (the silenced regions of DNA in the chromatin and are stained intensively by the DNA-binding dyes) by the RdDM pathway. In the first step of heterochromatin formation, RNAP IV couples with an RNA-dependent RNA polymerase (RDR2) to synthesize double-stranded RNA precursors from the single-stranded RNA transcripts of RNAP IV at all repeated loci. In the next step, a dicer-like protein (DLP3), an enzyme belonging to RNase III family, slices the double-stranded RNA precursors into 24 nts long siRNAs. These siRNAs are then methylated at their 3'-ends by a protein known as HUA-Enhancer 1 (HEN1) and finally, these methylated siRNAs complex with a protein known as Argonaute (AGO4) to form the silencing complex. Now the silencing complex performs methylation of the target regions on the chromosomes resulting in the formation of heterochromatin [17,31,32].

The gene silencing pathway involves in switching off genes. In the first step, they generate noncoding RNAs. The noncoding RNAs accomplish gene-silencing at transcriptional level (TGS) via the RdDM pathway. Thus, the RdDM pathway essentially requires two types of noncoding RNAs, (i.e.), one from the RNAP IV-dependent 24-nt siRNAs and the second one from the RNAP V-dependent intergenic noncoding (IGN) RNAs. They mediate cytosine methylation of complementary DNA sequences in all sequence contexts (**C**G, **C**HG and **C**HH, where H=A, C or T) by a *de novo* methyltransferase. Thus, the RdDM pathway involves in silencing of thousands of transposons, endogenous repeats, invading RNA viruses, transgenes and also some protein-coding genes [17 and references therein]. Therefore, analysis of these two different plant-specific RNAPs offers insight into their unique active sites and PR functions in gene silencing pathways in plants.

Figure 15 shows the MSA analysis of the elongation subunit of MSU RNAP IV. The template-binding and catalytic pairs are highlighted in yellow and the Zn-binding conserved Cs is highlighted in orange. The completely conserved $Mg^{2+}$-binding site is highlighted in light green. The polymerase region and the $Mg^{2+}$-binding site are highly conserved among them. The MSA analysis data shows that the catalytic and template-binding pairs in the polymerase/PR active sites are similar in the elongation subunits of RNAP III (-PRM[1]GPPNKKSIC[10]TTC[13]EGNFQNC[20]PGHY[24] GYLKL- *A. thaliana*) and RNAP IV (-SRL[1]GLPNPDSVC[10]RTC[13]GSKDRKVC[21]EGHF[25] G**V**INF - *A. thaliana*).

The distance conservations between the catalytic amino acid and first C of the $Zn^{2+}$-binding are also maintained in RNAP IV suggesting its possible origin from RNAP III rather than from RNAP II (Table 2). Furthermore, the invasive fungal pathogens use either a V or an I immediately after the template-binding pairs in their C1 elongation subunits of RNAP III (Fig. 12) as in RNAP IV (highlighted in red) further supporting their possible origin from PNAP III. The three invariant Ds in $Mg^{2+}$-binding regions, viz. -DxDxD- are completely conserved in all (Table 2).

Figure 16 shows the MSA analysis of the elongation subunits of the MSU RNAP V from various plant sources. The template-binding and catalytic pairs are highlighted in yellow and the Zn-binding conserved Cs is highlighted in orange. The completely conserved $Mg^{2+}$-binding site is highlighted in light green. The polymerase region and the $Mg^{2+}$-binding site are highly conserved in all. Unlike RNAP IV, some of them possess a conserved repeat in their CTD, a characteristic feature that is found only in the elongation subunits of RNAP II, but the sequence motifs are different. The distance conservations of the template binding pairs and the conserved Cs of the $Zn^{2+}$-binding sites are different in both the elongation subunits of RNAP II and RNAP V, e.g., -[62]R**K**V[1]KC[3]ETC[6]MANMAEC[13]PGHF[17]GYLELAK-

in RNAP II from *A. thaliana* and -$^{41}$S**QL**$^{1}$TNAFLGLPLEFGKC$^{15}$ESC$^{18}$GATEPDKC$^{2}$ $^{6}$EGHF$^{30}$GYIQLPVP- in RNAP V from *A. thaliana,* showing conservation only in the template-binding pairs (-FG-), but differ in their possible catalytic amino acids (-**K**V- in RNAP II and –**QL**- in RNAP V-)' Furthermore, the distance between the template binding pair and the catalytic pair is markedly different between RNAP II and RNAP V (Fig. 16 and Table 2). However, the $Zn^{2+}$ binding Cs are completely

CLUSTAL O (1.2.4) MSA of the elongation subunits of RNAP IV from various plant sources (Only polymerization/ PR sites and $Mg^{2+}$-binding sites are shown)

```
tr|A0A3P5Z8F9|A0A3P5Z8F9_BRACM      EAANQVTDSRLGLPNPDSICRSCGSKDRKVCEGHFGVINLKYPVINPYFLKEIATLLNRI      96
tr|A0A3P6CGM6|A0A3P6CGM6_BRAOL      EAANQVTDSRLGLPNPDSICRSCGSKDRKVCEGHFGVINLKYPVINPYFLKEIATLLNKI      96
tr|A0A6J0KXG2|A0A6J0KXG2_RAPSA      EAANQVTDSRLGLPNRDDSCKSCGSKDRKVCEGHFGVIKLSDPVINPHFLKEIATLLNKF      96
tr|A0A3P6DRW1|A0A3P6DRW1_BRAOL      EAANQVTDSWLGLPNRDDICKSCGNKDRKVCEGHFGVINLRYPVINPHLLKEVATLLNKF      96
tr|M4EFP9|M4EFP9_BRARP              EAANQVTDSRLGLPNRDDICKSCGSKDRKVCEGHFGVINLRYPVINPHLLKEVATLLNKF      96
tr|A0A1J3IW39|A0A1J3IW39_NOCCA      EAANQVTDSRLGLPNPESVCRSCGSKDRKVCEGHFGVINFKYSIINPYFVKEIAALLNKI      116
tr|A0A1J3GF20|A0A1J3GF20_NOCCA      EAANQVTDSRLGLPNPESVQRSCGSKDRKVCEGHFGVINFKYSIINPYFVKEIAALLNKI      119
sp|Q9LQ02.1|NRPD1_ARATH             EAPNQVTDSRLGLPNPDSVCRTCGSKDRKVCEGHFGVINFAYSIINPYFLKEVAALLNKI      96
tr|D7KU36|D7KU36_ARALL              EAPNQVTDSRLGLPNPDSICKTCGSKDRKVCEGHFGVINFQYSIINPYFLKEIAALLNKI      96
tr|V4MCZ2|V4MCZ2_EUTSA              EAPNQVSDSRLGLPNPDGICRACGSKDRKTCEGHFGVIKFQYSVINPYFLKEIAALLNRF      97
tr|A0A565AW69|A0A565AW69_9BRAS      EAANQVTDSRLGLPNPDSVCRTCGSKDRKVCEGHFGVIKFRHSIINPYFLKEIAALLNKI      96
tr|A0A087HEJ6|A0A087HEJ6_ARAAL      EAANQVTDSRLGLPNPDTICRTCGSKDRKGCEGHFGVIKFQNPILNPYFLKEVAALLNKI      96
                                    ** ***:** *****  :  *::**.**** ******** ::*   ::**:::**:*:***::

//
tr|A0A3P5Z8F9|A0A3P5Z8F9_BRACM      DDLQTGDSILRTLKDGDTVLMNRPPSIHQHSLIAMSVKVLPTTSVVSLNPICCLPFRGDF      449
tr|A0A3P6CGM6|A0A3P6CGM6_BRAOL      DDLQTGDSILRTLKDGDTVLMNRPPSIHQHSLIAMSVKVLPTTSVVSLNPICCLPFRGDF      449
tr|A0A6J0KXG2|A0A6J0KXG2_RAPSA      DDLQTGDSILRTLKDGDTVLMNRPPSIHQHSLIAMSVKLLPTASVVSLNPICCLPFRGDF      453
tr|A0A3P6DRW1|A0A3P6DRW1_BRAOL      DDLQTGDSILRTLKDGDTVLMNRPPSIHQHSLIAMSVKVLPTTSVVSLNPICCLPFRGDF      453
tr|M4EFP9|M4EFP9_BRARP              DDLQTGDSFLRTLKDGDTVLMNRPPSIHQHSLIAMSVKVLPTTSVVSLNPICCLPFRGDF      438
tr|A0A1J3IW39|A0A1J3IW39_NOCCA      DDLQTGDKIFRTLMDGDTVLMNRPPSIHQHSLIAMSVKVLPTTSVVSLNPICCLPFRGDF      471
tr|A0A1J3GF20|A0A1J3GF20_NOCCA      DDLQTGDKIFRTLMDGDTVLMNRPPSIHQHSLIAMSVKVLPTTSVVSLNPICCLPFRGDF      474
sp|Q9LQ02.1|NRPD1_ARATH             NDLQTGDKIFRSLMDGDTVLMNRPPSIHQHSLIAMTVRILPTTSVVSLNPICCLPFRGDF      448
tr|D7KU36|D7KU36_ARALL              SDLQTGDKIFRNLMDGDTVLMNRPPSIHQHSLIAMTVRVLPTTSVVSLNPICCLPFRGDF      448
tr|V4MCZ2|V4MCZ2_EUTSA              DDLQTGDNIFRTLRDGDTVLMNRPPSIHQHSLVAMSVRVLPTTSVVSLNPIICLPFRGDF      451
tr|A0A565AW69|A0A565AW69_9BRAS      DDLQNGDNIFRTLTDGDTVLMNRPPSIHQHSLIAMKVRVLPTTSVVSLNPICCLPFRGDF      449
tr|A0A087HEJ6|A0A087HEJ6_ARAAL      DDLQNGDQIFRTLTDGDTVLMNRPPSIHQHSLIAMSVRLLPTTSVVSLNPICCLPFRGDF      449
                                    .***.**.::*.* ******************:** *::***:******** ******** **

tr|A0A3P5Z8F9|A0A3P5Z8F9_BRACM      DGDCLHGYVPQSIQAKVELDELVALDKQLVNRQNGRNLLSLGQDSLTAAYIVSVETNCFL      509
tr|A0A3P6CGM6|A0A3P6CGM6_BRAOL      DGDCFHGYVPQSIQAKVELDELVALDKQLVNRQNGRNLLSLGQDSLTAAYIVNVETNCFL      509
tr|A0A6J0KXG2|A0A6J0KXG2_RAPSA      DGDCLHGYVPQSIQAKAELDELLGLDKQLVNRQNGRNLLALGQDSLSAAYILANVETNSFL      513
tr|A0A3P6DRW1|A0A3P6DRW1_BRAOL      DGDCLHGYVPQSIQAKAELDELVALDKQLVNRQNGRNLLSLGQDSLTAAYIVNVETNCFL      513
tr|M4EFP9|M4EFP9_BRARP              DGDCLHGYVPQSIQAKAELEELVALDKQLVNRQNGRNLLSLGQDSLTAAYIVNVETNCFL      498
tr|A0A1J3IW39|A0A1J3IW39_NOCCA      DGDCLHGYVPQSIQAKVELDELVALDKQLVNRQNGRNLLALGQDSLTGAYIVNVEKNCYL      531
tr|A0A1J3GF20|A0A1J3GF20_NOCCA      DGDCLHGYVPQSIQAKVELDELVALDKQLVNRQNGRNLLALGQDSLTGAYIVNVEKNCYL      534
sp|Q9LQ02.1|NRPD1_ARATH             DGDCLHGYVPQSIQAKVELDELVALDKQLINRQNGRNLLSLGQDSLTAAYLVNVEKNCYL      508
tr|D7KU36|D7KU36_ARALL              DGDCLHGYVPQSIQAKVELDELVALDKQLINRQNGRNLLSLGQDSLTAAYIVNVEKNCYL      508
tr|V4MCZ2|V4MCZ2_EUTSA              DGDCLHGYVPQSVQAKVELDELVALDKQFVNRQNGRNLLSLGQDSLTAAYIVNVEKNCFL      511
tr|A0A565AW69|A0A565AW69_9BRAS      DGDCLHGYVPQSVQARVELDELVALDKQLINRQNGRNLLSLGQDSLTAAYIVNVEKNCFL      509
tr|A0A087HEJ6|A0A087HEJ6_ARAAL      DGDCLHGYVPQSVQARVELNELVALDKQLINRQNGLNLLSLGQDSLTAAYIVNVEKNCYL      509
                                    ****:*******:**:.**:**:. .****::***** ***:******:.***..**.*.:*

// End of the elongation subunits of RNAP IV
tr|A0A3P5Z8F9|A0A3P5Z8F9_BRACM      LRNGTVQAGAI--------     1444
tr|A0A3P6CGM6|A0A3P6CGM6_BRAOL      LRNGTVQAGTI--------     1447
tr|A0A6J0KXG2|A0A6J0KXG2_RAPSA      LRNGTVKPGAI--------     1459
tr|A0A3P6DRW1|A0A3P6DRW1_BRAOL      LRNGTVQPGAV--------     1459
tr|M4EFP9|M4EFP9_BRARP              LRNGTVQPGAI--------     1443
tr|A0A1J3IW39|A0A1J3IW39_NOCCA      LNRGTM-------------     1491
tr|A0A1J3GF20|A0A1J3GF20_NOCCA      LNRGMV-------------     1495
sp|Q9LQ02.1|NRPD1_ARATH             LKNGTLESGGFSENP----     1453
tr|D7KU36|D7KU36_ARALL              LKNGTHQPGRLSENPQTVK     1456
tr|V4MCZ2|V4MCZ2_EUTSA              LNPKPVKENPKTR------     1465
tr|A0A565AW69|A0A565AW69_9BRAS      LKRGTVKPVTL--------     1457
tr|A0A087HEJ6|A0A087HEJ6_ARAAL      LKNGTVPPVTL--------     1457
                                    *.
```

**Fig. 15. MSA of the elongation subunits of the RNAP IV from various plant sources**

*A0A3P5Z8F9_BRACM Brassica campestris*
*A0A3P6CGM6_BRAOL Brassica oleracea*
*A0A6J0KXG2_RAPSA Raphanus sativus*
*A0A3P6DRW1_BRAOL Brassica oleracea*
*M4EFP9_BRARP Brassica rapa subsp. Pekinensis*
*A0A1J3GF20_NOCCA Noccaea caerulescens*
*A0A1J3IW39_NOCCA Noccaea caerulescens*
*Q9LQ02|NRPD1_ARATH Arabidopsis thaliana*
*D7KU36_ARALL Arabidopsis lyrata subsp. Lyrata*
*V4MCZ2_EUTSA Eutrema salsugineum*
*A0A565AW69_9BRAS Arabis nemorensis*
*A0A087HEJ6_ARAAL Arabis alpine*

CLUSTAL O (1.2.4) MSA of the elongation subunits of RNAP V from various plant sources (Only polymerization/PR and the $Mg^{2+}$-binding regions are shown)

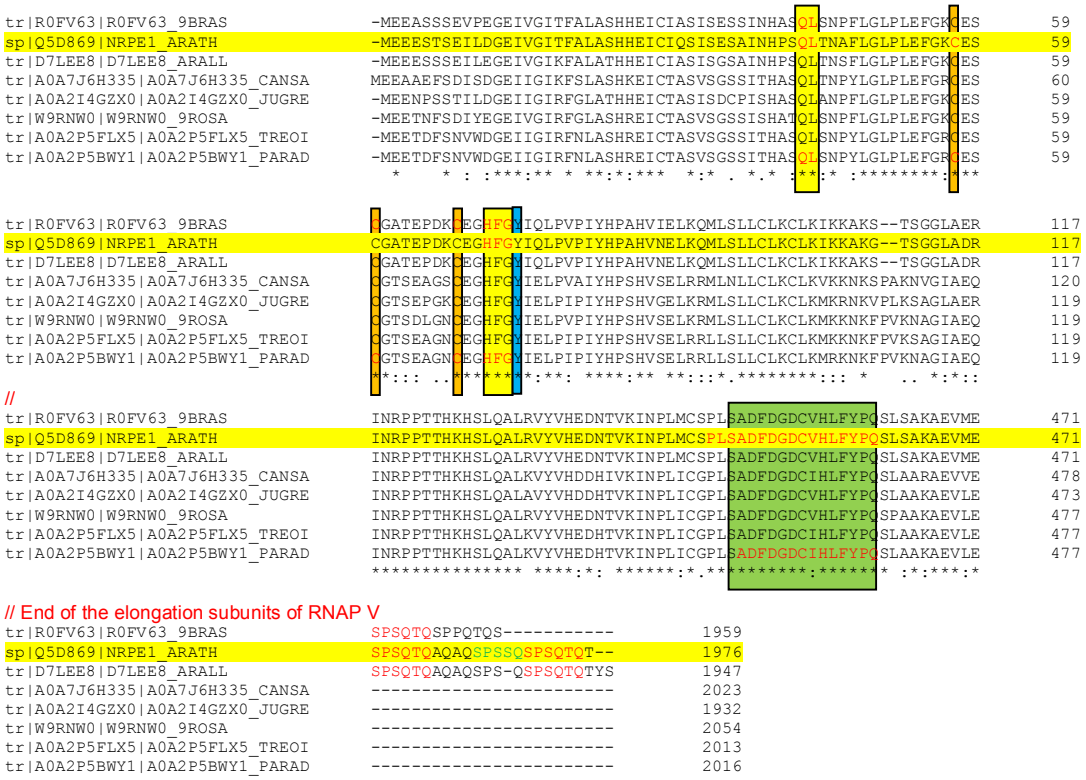



**Fig. 16. MSA of the elongation subunits of the RNAP V from various plant sources**

*R0FV63_9BRAS Capsella rubella*
*D7LEE8_ARALL Arabidopsis lyrata*
*A0A2I4GZX0_JUGRE Juglans regia*
*A0A2P5FLX5_TREOI Trema orientale*
*Q5D869|NRPE1_ARATH Arabidopsis thaliana*
*A0A7J6H335_CANSA Cannabis sativa*
*W9RNW0_9ROSA Morus notabilis*
*A0A2P5BWY1_PARAD Parasponia andersonii*

conserved within the polymerase region, but exhibits different distance conservations in both the types of RNAPs (Table 2). The three invariant Ds in $Mg^{2+}$-binding sites, viz. -DxDxD- are completely conserved in RNAPs I, II, III, IV and V, e.g., -YNADFDGDEINVHFPQ$^{146}$ – in RNAP I from *A. thaliana*; -YNADFDGDEMNMHVPQ$^{507}$- in RNAP II from *A. thaliana,* -YNADFDGDEMNMHVPQ$^{511}$- RNAP III from *A. thaliana,* -FRGDFDGDCLHGYVPQ$^{459}$- in RNAP IV from *A. thaliana* and -LSADFDGDCVHLFYPQ$^{461}$- in RNAP V from *A. thaliana*, and but showing variations in their flanking regions, (i.e.), RNAP IV and V flanking regions are different from other three RNAPs (Table 2).

Table 2 shows a consolidated account of the proposed polymerase, PR and $Mg^{2+}$-binding sites in all the seven (I-VII) MSU RNAPs, both from prokaryotes and eukaryotes. It is clear from Table 2 that the $Zn^{2+}$-binding invariant Cs are conserved in all the seven MSU RNAPs within the polymerase active site region itself, suggesting a possible Zn-mediated excision of mismatched nucleotides during PR activity. The template-binding (-YG-) and the catalytic pairs (-RS/T-) are almost identical in the prokaryotic and prokaryotic-type RNAPs from eubacteria and plant chloroplasts. The highly conserved 3 Cys residues within the polymerase active site is also found in both the bacterial and chloroplast RNAPs and shown to bind a zinc atom by crystallography studies [12]. They also maintain similar distance conservation in both. Possible proton acceptor for polymerase and PR reactions are shown in red and dark blue, respectively.

Among the 3 common RNAPs found in all eukaryotes (I-III), RNAP II is remarkably

conserved in the organisms, viz. from yeasts, higher fungi, plants and animals. The plant-specific RNAPs IV and V exhibit similarities to RNAP III, i.e., having a similar polymerase active site with the integrated $Zn^{2+}$-binding site and with more or less similar distance conservation between the catalytic amino acid and the first C in the zinc-binding site. These similarities suggest that the plant-specific RNAPs IV and V might have evolved from RNAP III, as all three RNAPs are localized in the nucleus and involve in transcription of low molecular weight RNAs. RNAP IV transcription is found to be the most error-prone [30] and it also uses unusually a Val

**Table 2. Summary of the proposed polymerase/PR active sites and MBS^ in MSU RNAPs***

| TYPE and SOURCE | POLYMERASE/PR ACTIVE SITES* | $Mg^{2+}$-BINDING SITE |
|---|---|---|
| **PROKARYOTES** | | |
| **a. MSU RNAP from Bacteria** | | |
| Eubacteria (*E. coli*) | -$^{883}$VRS$^{1}$VVSC$^{5}$DTDFGVC$^{12}$AHC$^{15}$YGRDLARG-(**β'**) | -AYNADFDGDQMAVHVPL$^{472}$- (**β'**) |
| **b. MSU RNAP from Plant Chloroplasts (Prokaryotic-type)** | | |
| Chloroplasts (*A. thaliana*) | -$^{287}$IRT$^{1}$PFTC$^{5}$RSTSWIC$^{12}$RLC$^{15}$YGRSPTHG-(**β''**) | -GFNADFDGDQMAVHVPL$^{501}$- (**β'**) |
| Chloroplasts (*Z. mays*) | -$^{290}$IRT$^{1}$PFTC$^{5}$RSTSWIC$^{12}$QLC$^{15}$YGRSPTHG-(**β''**) | -GFNADFDGDQMAVHLPL$^{501}$- (**β'**) |
| **EUKARYOTES** | | |
| **A. RNAP I (A1) from Fungal, Plant and Animal Sources** | | |
| Yeast (*Sc*) A1 | -$^{58}$LRN$^{1}$LC$^{3}$STCGLDEKFC$^{13}$PGHQGHIELPV- | -AYNADFDGDEMNMHFPQ$^{639}$- |
| Higher fungi (*Pc*) A1 | -$^{60}$DHV$^{1}$C$^{2}$TTCRQNSFTC$^{12}$TGHPGHIELPV- | -TYNADFDGDEMNMHFPQ$^{633}$- |
| Plants (*At*) A1 | -$^{75}$DKQ$^{1}$AC$^{3}$NSCGQLKLAC$^{13}$PGHCGHIELVFPI- | -TYNADFDGDEINVHFPQ$^{146}$- |
| Plants (*Ah*) A1 | -$^{60}$EKL$^{1}$PC$^{3}$KTCGQLYHLC$^{13}$PGHFGRIELVSPV- | -TYNADFDGDEINVHFPQ$^{120}$- |
| Animals (Human) | -$^{60}$SKE$^{1}$VC$^{3}$STCVQDFSNC$^{13}$SGHLGHIELPL- | -AYNADFDGDEMNAHFPQ$^{600}$- |
| **B. RNAP II (B1) from Fungal, Plant and Animal Sources** | | |
| Yeast (*Sc*) B1 | -$^{62}$DRN$^{1}$LKC$^{4}$QTCQEGMNEC$^{14}$PGHFGHIDLAK- | -PYNADFDGDEMNLHVPQ$^{493}$- |
| Higher fungi (*Nc*) B1 | -$^{64}$DRQ$^{1}$FKC$^{4}$KTCGENMSEC$^{14}$PGHFGHIELAR- | -PYNADFDGDEMNLHVPQ$^{507}$- |
| Plants (*At*) B1 | -$^{62}$RKV$^{1}$KC$^{3}$ETCMANMAEC$^{13}$PGHFGYLELAK- | -PYNADFDGDEMNMHVPQ$^{506}$- |
| Plants (Wheat) B1 | -$^{62}$RRT$^{1}$KC$^{3}$ETCMAGMAEC$^{13}$PGHFGHLELAK- | -PYNADFDGDEMNMHVPQ$^{507}$- |
| Animals (Human) B1 | -$^{66}$ERT$^{1}$GRC$^{4}$QTCAGNMTEC$^{14}$PGHFGHIELAK- | -PYNADFDGDEMNLHLPQ$^{507}$- |
| **C. RNAP III (C1) from Fungal, Plant and Animal Sources** | | |
| Yeast (*Sc*) C1 | -$^{55}$DPKM$^{1}$*GVSSSSLEC*$^{10}$ATCHGNLASC$^{20}$HGHFGHLKLAL- | -PYNADFDGDEMNLHVPQ$^{523}$- |
| Higher fungi (*Um*) C1 | –$^{56}$DRRL$^{1}$*GVSDKNSLC*$^{10}$ETCHLKMADC$^{20}$VGHYGYIKLVL- | -PYNADFDGDEMNMHVPQ$^{510}$- |
| Plants (*At*) C1 | -$^{63}$DPRM$^{1}$GPPNKKSIC$^{10}$TTCEGNFQNC$^{20}$PGHYGYLKLDL- | -PYNADFDGDEMNMHVPQ$^{511}$- |
| Plants (Rice) C1 | -$^{64}$DTRM$^{1}$GAANKLGEC$^{10}$STCHGSFAEC$^{20}$PGHFGYLKLAL- | -PYNADFDGDEMNLHVPQ$^{499}$- |
| Animals (Human) C1 | -$^{57}$DHRM$^{1}$GTSEKDRPC$^{10}$ETCGKNLADC$^{20}$LGHYGYIDLEL- | -PYNADFDGDEMNLHLPQ$^{510}$- |
| **D. RNAPs IV AND V - Plant-specific** | | |
| *A. thaliana* IV | -$^{45}$SRL$^{1}$GLPNPDSVC$^{10}$RTCGSKDRKVC$^{21}$EGHF$^{25}$GVINFAY- | -PFRGDFDGDCLHGYVPQ$^{459}$- |
| *A. thaliana* V | -$^{30}$SQL$^{1}$TNAFLGLPLEFGKC$^{15}$ESCGATEPDKC$^{26}$EGHF$^{30}$GYIQLPV- | -PLSADFDGDCVHLFYPQ$^{461}$- |

**Proposed,* ^MBS, Metal-binding site; *unusual type of template binding amino acids is highlighted in light green.*
*Sc, Saccharomyces cerevisiae; Pc, Penicillium chrysogenum; At, Arabidopsis thaliana;*
*Ah, A. hypogaea; Um, Ustilago maydis. Nc, Neurospora crassa*
*The $Mg^{2+}$ ions which positions the NTP at the polymerization active site is completely conserved*

residue immediately after the template-binding pair (Table 2). When compared to RNAPs III and V, RNAP IV transcripts are ~ 30 to 40 nts in length, long enough for formation of 24-nt SiRNAs. Whereas all the 6 RNAPs have a proton acceptor amino acid (R/H/Y) immediately downstream of the template-binding pairs, only the RNAP IV uses a different amino acid (Val) at that position, possibly explaining its error–prone nature [30] (Table 2). It is interesting to note, that the RNAP III from invasive fungal pathogens also use either a Val or an Ile, immediately after the template-binding pair (Fig. 12), similar to RNAP IV. Kostyuk et al. [33] shown that in T7 RNA polymerase, a single amino acid substitution at the template binding triad –YG**S**- (**S**$^{641}$→A) lost its ability to discriminate NTPs/dNTPs and was able to synthesize DNA, suggesting that the downstream amino acid at the template-binding pair plays an important role in nucleotide selection, and hence suggesting the possible error-prone nature of RNAP IV.

It is interesting to note that the invariant Asn (**N**) in the $Mg^{2+}$-binding site –**N**ADFDFD-, which is shown to involve in nucleotide selection, is completely conserved in all MSU RNAPs except in the plant-specific RNAPs IV and V, where an -**R** and -**S** are used, respectively (Table 2).

Table 3 shows the proposed PR sites and their distance conservations between the C residues in Zn-binding motif of the elongation subunits of prokaryotes (β'), eukaryotes (A1, B1 and C1), plant chloroplasts (β'') and plant-specific MSU RNAPs. It is clear from Table 3 that the distance between the first and the last C is always maintained at 11/12, suggesting a Zn-binding site in all the MSU RNAPs and therefore, a Zn-mediated excision of the mismatched nucleotide during transcription is proposed. In all the cases, the PR site is embedded within the active site of the respective MSU RNAPs. Furthermore, the template binding pair is immediately followed by a basic amino acid R/H except in the plant-specific RNAPs IV and V.

Table 4 shows the uncommon template-binding pairs like –SG-, –PG-, -MG-, -AG- that are observed in the elongation subunits of RNAP I and RNAP III. Uncommon template-binding pairs are seen in the RNAP I of yeasts and higher fungi, where they are found to be either human pathogens or could not ferment sugars. Uncommon template-binding pairs and its immediate downstream amino acid are also observed in the RNAP III of animal and human pathogens. These results suggest that pathogens, parasites and organisms which could not ferment sugars are adapted to different types of template-binding pairs. It is not clear whether such differences offer any advantages to these organisms for transcription in their hosts.

**Table 3. Distance conservations of Cs in ZBS* in the elongation subunits of MSU RNAPs**

| Organism | Proposed Polymerase/PR Active Sites | Distance Conservations of Cs |
|---|---|---|
| ***1. Prok-MSU RNAP (Ec)*** | -$^{883}$V**R**S$^{1}$VVS**C**$^{5}$DTDFGV**C**$^{12}$A**H**C$^{15}$**Y**$^{16}$**G**R- **(β')** | (C5+C12+C15 = **C1 to C3 = 11**) |
| **2a. Chloroplasts (*At*)** | -$^{287}$I**R**T$^{1}$PFT**C**$^{5}$RSTSWI**C**$^{12}$**R**LC$^{15}$**YG**R- **(β'')** | (C5+C12+C15 = **C1 to C3 = 11**) |
| **2b. Chloroplast (*Zm*)** | -$^{292}$I**R**T$^{1}$PFT**C**$^{5}$RSTSWI**C**$^{12}$**Q**LC$^{15}$**Y**$^{16}$**G**R- **(β'')** | (C5+C12+C15 = **C1 to C3 = 11**) |
| ***3. Euk-* Pol I-A1 (*Sc*)** | -$^{58}$L**R**N$^{1}$LC$^{3}$ST**C**$^{6}$GLDEKF**C**$^{13}$PG**H**Q$^{17}$**G**H- | (C3+C6+C13 = **C1 to C3 = 11**) |
| ***4. Euk-* Pol II-B1 (*Sc*)** | -$^{62}$D**R**N$^{1}$LK**C**$^{4}$QT**C**$^{7}$QEGMNE**C**$^{14}$PG**H**F$^{18}$**G**H- | (C4+C7+C14 = **C1 to C3 = 11**) |
| ***5. Euk-* Pol III-C1 (*Sc*)** | -$^{56}$P**K**M$^{1}$GVSSSSLE**C**$^{10}$AT**C**$^{13}$HGNLAS**C**$^{20}$HG**H**F$^{24}$**G**H- | (C10+C13+C20 = **C1 to C3 = 11**) |
| ***6. Euk- Pol* IV (*At*)** | -$^{45}$S**R**L$^{1}$GLPNPDSV**C**$^{10}$RT**C**$^{13}$GSKDRKV**C**$^{21}$EG**H**F$^{25}$**G**V- | (C10+C13+C21 = **C1 to C3 = 12**) |
| ***7. Euk- Pol V (At)*** | -$^{30}$S**Q**L$^{1}$TNAFLGLPLEFGK**C**$^{15}$ES**C**$^{18}$GATEPDK**C**$^{26}$EG**H**F$^{30}$**G**Y- | (C15+C18+C26 = **C1 to C3 = 12**) |

*Ec, E. coli; Zm, Z. mays; Sc, Saccharomyces cerevisiae; At, Arabidopsis thaliana*

*ZBS, Zinc-binding site in the proposed PR site within the polymerization active site is shown

**Table 4. Uncommon template-binding pairs in the MSU RNAPs I and III**

| RNAP/Organism | Nature of the Organism | Catalytic and Template-Binding Pairs* | MBS^ |
|---|---|---|---|
| **RNAP I (Elongation subunit, A1)** | | | |
| **a. From Yeasts** | | | |
| *Tortispora caseinolytica* | Unable to ferment sugars | -$^{59}$LRHPC$^{3}$ATCRLDERFC$^{13}$PGHSGHIEL- | Identical |
| *Yarrowia lipolytica* | Unable to ferment sugars | -$^{59}$LRNNC$^{3}$ATCNLDNRFC$^{13}$QGHPGHIEL- | Identical |
| *Candida lipolytica* | Human pathogen | -$^{59}$LRNNC$^{3}$ATCNLDNRFC$^{13}$QGHPGHIEL- | Identical |
| *Spathaspora passalidarum* | Unable to ferment sugars | -$^{59}$LRNVC$^{3}$TTCGLDEKFC$^{13}$PGHMGHIEL- | Identical |
| *Candida albicans* | Human pathogen | -$^{59}$LRNLC$^{3}$TTCGLDEKFC$^{13}$PGHMGHIEL- | Identical |
| **b. From Higher Fungi** | | | |
| *Talaromyces marneffei* | Human pathogen | -$^{60}$*D*HVC$^{2}$TTCRASSWSC$^{12}$PGHPGHIEL- | Identical |
| *Talaromyces cellulolyticus* | Cellulolytic fungus | -$^{60}$*D*HVC$^{2}$TTCRASSWSC$^{12}$PGHPGHIEL- | Identical |
| *Ajellomyces capsulatus#* | Human pathogen | -$^{60}$*D*HIC$^{2}$TTCRLNSWSC$^{12}$NGHAGHIEL- | Identical |
| *Blastomyces dermatitidis* | Human pathogen | -$^{60}$*D*HVC$^{2}$TTCRQNSFTC$^{12}$TGHPGHIEL- | Identical |
| **RNAP III (Elongation subunit, C1)** | | | |
| **a. From Human and Animal Parasites** | | | |
| *Encephalitozoon cuniculi* | Animal parasite | –$^{55}$DLRL$^{1}$GVGNKKDKC$^{10}$ATC$^{13}$GEGLATC$^{20}$IGHFGEVRL- | Identical |
| *Vittaforma corneae* | Animal parasite | –$^{55}$DLRL$^{1}$GVSTKSGIC$^{10}$STCKENIQNC$^{20}$AGHFGQIQL- | Identical |
| *Plasmodium falciparum* | Human parasite | -$^{76}$DLKL$^{1}$GAHKSNSVC$^{10}$ETCNKKLINC$^{20}$SGHFGYIEL- | Identical |
| *Neospora caninum* | Animal parasite | -$^{180}$DLRL$^{1}$GPNKSDSRC$^{10}$QTCGHTLLQC$^{20}$TGHWGYMDL- | Identical |
| *Toxoplasma gondii* (M) | Animal & Human parasite | -$^{153}$DLRL$^{1}$GPNKSDSRC$^{10}$QTCGHTLLQC$^{20}$TGHWGYMDL- | Identical |
| *Toxoplasma gondii* (O ) | Animal & Human parasite | -$^{154}$DLRL$^{1}$GPNKSDSRC$^{10}$QTCGHTLLQC$^{20}$TGHWGYMDL- | Identical |

**Proposed; ^MBS, $Mg^{2+}$-binding site with 3 invariant Ds as –DxDxD-*

## Uncommon template-binding pairs in:

### RNAP I, A1 subunits– From Yeasts

***Tortispora caseinolytica*** occurs in the rotting tissues of opuntias cactus in the Sonoran desert. It does not ferment any sugar, but could assimilate only a limited number of carbon compounds. ***Yarrowia lipolytica*** is characterized by its inability to ferment sugars but possesses unique physiological capabilities to utilize polyalcohols, organic acids, and long-chain hydrocarbons, crude oils, *n*-paraffins and hydrocarbons. ***Yarrowia lipolytica lyrata*** *also known as* ***Candida lipolytica*** is a human pathogen but rarely infect humans. ***Spathaspora passalidarum*** (Debaryomycetaceae) is a wood-boring beetle associated fungus and one of a few yeasts known to efficiently ferment and metabolise xylose, a major component of plant cell walls. This rare ability enables this for the production of biofuel from plant materials. ***Candida albicans*** is the most prevalent cause of fungal infections in humans including urinary tract infections.

### RNAP I, A1 subunits– From Higher Fungi

***Talaromyces (Penicillium) marneffei)*** *is a human pathogen.* Patients are presented with fever, altered mental status, headaches, facial nerve palsy, seizures, vision loss and mainly affects immuno-compromised patients like HIV/AIDS. ***Talaromyces cellulolyticus*** is a cellulolytic fungus which could utilize the complex carbohydrate. ***Ajellomyces capsulatus,*** also known as *Histoplasma capsulatum* a thermal dimorphic fungus that causes histoplasmosis, a potentially fatal disease of the lungs. ***Blastomyces dermatitidis*** is the causal agent of blastomycosis, an invasive and often serious fungal infection.

### RNAP III, C1 subunits- From Animal and Human Parasites

***Encephalitozoon cuniculi*** is a microsporidial parasite (intracellular animal pathogen) that causes renal and central nervous system diseases in farmed rabbits.

***Vittaforma corneae*** is a microsporidial pathogen cause keratitis and keratoconjuctivitis, an. obligate intracellular parasite adapted to parasitic life in a wide range of eukaryotic organisms. ***Plasmodium falciparum*** is a unicellular protozoan parasite of humans, and the deadliest species of Plasmodium that causes malaria in humans. ***Neospora caninum*** is a coccidian parasite of animals. It is a major pathogen of cattle and dogs. ***Toxoplasma gondii*** is one of the world's most common parasites and is considered to be a leading cause of death attributed to food-borne illness. It is a protozoan, single celled, obligate intracellular parasite that infects most species of warm-blooded animals, including humans, and causes the disease toxoplasmosis.

## 5. A UNIFIED MECHANISM OF ACTION FOR THE PR EXONUCLEASES OF MSU RNAPS

Unlike the PR functions in DNA polymerases, the PR function in MSU RNAPs is not very well understood. In an attempt to explain the PR exonuclease activity of MSU RNAPs, a unified mechanism of action is proposed in this communication, based on the completely conserved $Zn^{2+}$-binding site embedded within the polymerase active site region in all the 7 known MSU RNAPs from prokaryotes and eukaryotes. It has been well established that zinc acts as the cofactor for >450 enzymes and proteins, where it plays both structural and catalytic roles. It is interesting to note that zinc is also an integral component of both prokaryotic and eukaryotic DNA and RNA polymerases and is found to be essential for their function [34,35]. As the Zn-mediated reactions are exceedingly faster and efficient, many enzymes use $Zn^{2+}$ as the catalytic metal ion. For example, the mechanisms of several zinc metalloenzymes have been proposed to be facilitated by the formation of the highly reactive zinc-hydroxide. Palanivelu [1,7] has shown that $Zn^{2+}$ excise mismatched bases/ nucleotides in DNA and RNA polymerases as well as in RNA modifying enzymes. For example, PR functions in prokaryotic DNAPs like DNAP I, II, III and DNAP X and RdRps of SARS-CoV-2 are proposed to be mediated by Zn-mediated hydrolysis [36].

Involvement of $Zn^{2+}$ is in the PR function of the MSU RNAPs is arrived at based on two important findings. For example, the X-ray crystallographic analysis [12] and MSA data have shown that the MSU RNAPs harbour a Zn-binding site within the polymerase region itself. Therefore, Sydow and Cramer [11] suggested that the MSU RNAPS could use the same catalytic region for PR function too. In fact, Zenkin et al [16] studies on a prokaryotic MSU RNAP from the thermophilic bacterium, *Thermus thermophilus,* strongly supports their views. They have shown that when a wrong nucleotide is incorporated, the *T. thermophilus* RNAP stalls and moved one step backwards and made a cleavage at the penultimate base, resulting in the removal of a dimer which includes the wrongly added nucleotide. Subsequently, the polymerase resumed polymerization with the correct nucleotide inserted into the polymerization site. They further suggested that the terminal RNA nucleotide mismatch itself could play an active role in RNA PR activity. Furthermore, the first X-ray crystallographic analysis of a eukaryotic RNAP II complex from *S. cerevisiae* with its transcription factor and the elongation factor TFIIS, supported the idea that the polymerase has a 'tunable' active site that switches between mRNA synthesis and repair [10]. These results further strengthen the model of a bifunctional active site in MSU RNAPs [37]. Moreover, a 2.8 Å difference Fourier map revealed the presence of two metal ions at the active site of the yeast RNAP II, a persistently bound metal ion (metal A) and a mobile metal ion (metal B) suggesting that the metal ion A possibly participating in regular polymerization reactions and the mobile metal ion B in the PR function, only when a wrong nucleotide is inserted [38]. In addition to, the SDM experiments from both prokaryotic and eukaryotic MSU RNAPs have proved that the conserved Cs are not only essential for the enzyme activity but also lethal to the organism [13,25] suggesting also a structural role for the $Zn^{2+}$ in maintaining the polymerase and PR functions tightly coordinated. In support of these findings, all the 7 MSU RNAPs from prokaryotes and eukaryotes possess a completely conserved 3 Cys residues (the $Zn^{2+}$-binding site) and an invariant H (as proton acceptor) within the polymerase active site region itself. Only in chloroplast MSU RNAP, a Q or an R could possibly do the job as there was no invariant H at the expected position as in others. Based on these findings, a unified PR mechanism is proposed for these MSU RNAPs (Fig. 17). A schematic diagramme (Fig. 2A) shows the mechanism of action of the polymerase and PR exonuclease activities in prokaryotes (*E. coli*). Proposed steps for the mechanism of action PR exonuclease activity of eukaryotic (*S. cerevisiae*) MSU RNAP II is shown in Fig. 17.

Some amino acid(s) participate in sugar selection and some involve in base selection, and are

mostly non-overlapping. The N residue in the completely conserved $^{479}$**N**ADFDGD of the $Mg^{2+}$-binding site in all major RNAPs plays a crucial role in nucleotide discriminations. Functional *in vitro* analysis demonstrated that the substitutions of the corresponding **N**$^{458}$ residue in the prokaryotic (*E. coli*) elongation subunit β′ not only led to the loss of discrimination between NTP and dNTP substrates but also led to defects in RNA chain extension [19,39]. It is interesting to note that substitution of the corresponding amino acid in yeast elongation subunit, Rpb1 (**N**$^{479}$ →Y) is lethal in the same sequence motif $^{479}$**N**ADFDGD found in eukaryotes [40]. They suggested that the absolutely conserved N in both the cases could discriminate the NTP from dNTP by recognizing the 2'-OH of the ribose and suggested that the crucial N could interact with both the 2'-OH as well 3'-OH [40,41]. A similar mechanism should be operating in all other 6 MSU RNAPs as the PR site is integrated into the polymerase active site itself in all (Table 2). All of them use a basic amino acid R/K/H as the proton acceptor except the plant-specific RNAP V where a Q is found (the –QL- diad is found in the MSP RNAPs of chloroplasts and RNAP V).

## 5.1 Involvement of Additional Proteins for PR Functions

It was proposed that the PR activity in prokaryotes may use additional protein like GreA for excising the mismatches during transcription [36]. However, further insights into the participation of any extraneous protein factors for PR activity were provided by Zenkin et al [16]. They have shown that the RNAP efficiently cleaved the penultimate (P2) phosphodiester bond, but not the P1 (ultimate phosphodiester bond), suggesting that MSU RNAP backtracked (sliding backwards) by 1 base pair relative to the pre-translocated state and cleaved the penultimate phosphodiester bond. Furthermore, they have found that the selective removal of mismatched residues during transcription did not require GreA in *T. aquaticus* RNAP, proving that the PR activity is independent of other factors. They found that the cleavage factors are not also essential *in vivo*. In other words, their findings show that the mRNA itself could correct error(s) that might occur during its own synthesis. Similar findings were also reported for the eukaryotic RNAPs. For example, Khun et al [18] found that the conserved polymerase active site of RNAP I was capable of RNA cleavage in the absence of cleavage stimulatory factors. Similar observations were made for RNAP II and RNAP III also, (i.e.), that the hydrolytic activity is intrinsic to RNAPs II and III and factor-independent [27,42]. The intrinsic cleavage activity was stimulated greatly by mildly basic pHs and divalent metal ions. After the stalled, nascent transcript was cleaved by the intrinsic PR activity, they resume elongation as usual [27,42]. Therefore, during transcription elongation, a hydrolytic reaction stimulated by misincorporated nucleotides proofreads the misincorporation events and thus, serves as an intrinsic mechanism of transcription fidelity. The terminal mismatch nucleotide itself plays an active role by stimulating the repair reaction. Thus, the MSU RNAPs carryout both these functions using the same active site region. However, the transcription elongation factor TFIIs of RNAP II and its equivalent factors in RNAP I and RNAP III play a crucial role not only in the elongation process but also switching between polymerization and cleavage modes when mismatch occurs [11,16,28,43].

The release of dinucleotides, and larger oligonucleotides during PR activity has been detected by different workers [44,45]. When the RNAP II progress is blocked by mismatche(s), it is resolved by temporary backtracking of the RNAP and deleting of the mismatch in a transcription elongation factor S-II (TFIIS)-dependent or independent manner. Dinucleotides tend to originate from SII-*independent* mechanism, whereas 7–14 base products were observed from SII-dependent mechanism [44]. By the backtracking mechanism the enzyme retreats on the template resulting in the extruding of the 3'-end of the RNA and reaching the penultimate nucleotide for excision [45].

*Steps 1 and 2: The mismatch induces the PR reaction. The PR exonuclease frays the mismatched nucleotide from the DNA template and backtracks to the penultimate nucleotide. The PR exonuclease active site initiates proton transfer from the water-bound $Zn^{2+}$ with the simultaneous nucleophilic attack on the susceptible phosphodiester bond by the highly reactive Zn-hydroxide.*

*Step 3 and 4: The wrongly added nucleotide is excised along with the penultimate nucleotide as a dimer and the polymerase resume synthesis from penultimate nucleotide with the correct nucleotide.*

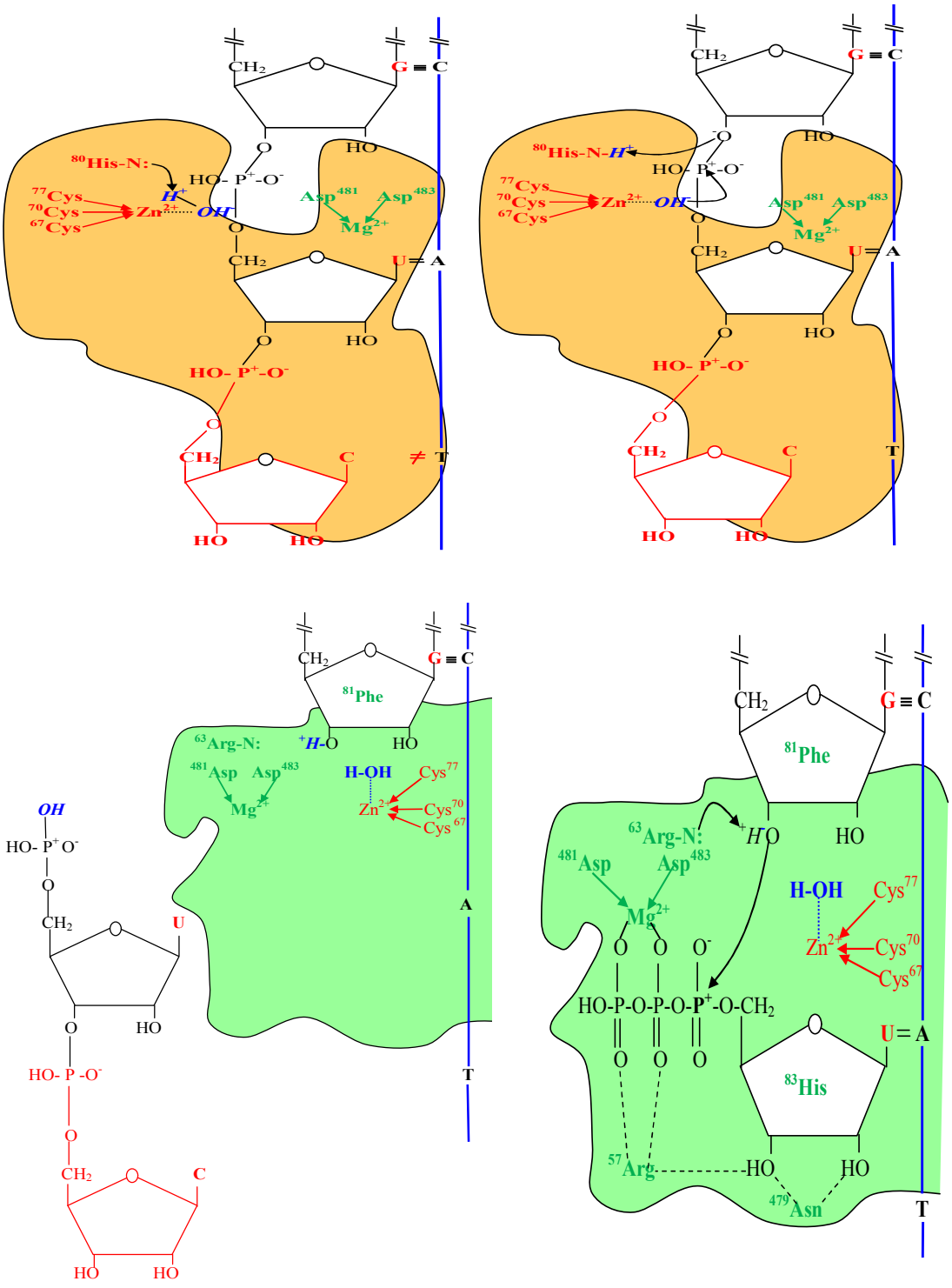



**Fig. 17. Steps (1-4) Involved in the Proposed Mechanism of Action of PR Function in MSU RNAPs II (*S. cerevisiae*)**

## 6. CONCLUSION

RNAPs play a crucial role in gene expression, where the genetic blueprint on DNA is copied into RNA. Some of these RNAPs are SSU types (e.g., viral, nuclear-encoded mitochondrial and chloroplast RNAPs) and others are MSU types. Prokaryotic, prokaryotic-type (chloroplast) and all the eukaryotic RNAPs are MSU types. These RNAPs rarely make mistakes (~$10^{-5}$) during the transcription process. As some of these mistakes could drastically affect the growth and very survival of the organisms, the RNAPs correct these mistakes by an intrinsic PR mechanism. Unlike in DNA polymerases, in the MSU RNAPs the PR mechanism is found to be integrated into the polymerase active site itself. This PR activity is proposed to be accomplished by a $Zn^{2+}$-mediated excision by the integrated $Zn^{2+}$-binding site in these MSU RNAPS. Thus, the discovery of an intrinsic mechanism in self-correcting RNA transcripts fulfils a missing link in molecular evolution.

## ACKNOWLEDGEMENTS

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

## COMPETING INTERESTS

Author has declared that no competing interests exist.

## REFERENCES

1. Palanivelu P. Eukaryotic Multi-subunit DNA dependent RNA Polymerases: An Insight into their Active Sites and Catalytic Mechanism. In: Advances and Trends in Biotechnology and Genetics. Sciencedomain International Book Publishers, UK. 2020;1:1-38. DOI: 10.9734/bpi/atbg/v1
2. Anikin M, Molodtsov V, Temiakov D, McAllister WT. Transcript slippage and recoding. In: Atkins JF, Gesteland RF, Bujnicki JM. editors. Recoding: Expansion of decoding rules enriches gene expression. 24th ed. Springer, New York; 2010.
3. Sahin U, Kariko K, Tũreci Ö. mRNA-based therapeutics- Developing a new class of drugs. Nat Rev Drug Discov. 2014;13:759–780.
4. Damase TR, Sukhovershin R, Boada C, Taraballi F, Pettigrew RI, John P. Cooke JP. The Limitless Future of RNA Therapeutics. Front. Bioeng. Biotechnol. 2021; 9: 1-24.
5. Conry RM, LoBuglio AF, Wright M, Sumerel L, Pike MJ. Characterization of a messenger RNA polynucleotide vaccine vector. Cancer Res.1995;55:1397–1400.
6. Pardi N, Hogan MJ, Porter FW, Weissman D. mRNA vaccines-a new era in vaccinology. Nat Rev Drug Discov. 2018;17:261–279.
7. Palanivelu P. DNA polymerases – An insight into their active sites and mechanism of action, In: Recent Advances in Biological Research, SCIENCEDOMAIN International Book Publishers, UK. 2019;1:1-39. DOI: 10.9734/bpi/rabr/v1
8. Palanivelu P. Multisubunit RNA Polymerases of Bacteria - An insight into their active sites and catalytic mechanism. Indian J Sci Technol. 2018;11:1-37.
9. Palanivelu P. Active Sites of the Multi-subunit RNA Polymerases of Eubacteria and Chloroplasts are Similar in Structure and Function: Recent Perspectives. In: Current Research Trends in Biological Science, SCIENCEDOMAIN International Book Publishers, UK. 2020;2:26-61. DOI: 10.9734/bpi/crtbs/v2
10. Cramer P RNA polymerase II structure: from core to functional complexes. Curr Opin Genet Dev. 2004;14:218-226.
11. Sydow JF, Cramer P. RNA polymerase fidelity and transcriptional proofreading. Curr Opin Struct Biol. 2009;19:732–739.
12. Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. Cell. 1999;98:811-824.
13. Zaychikov E, Martin E. Denissova L, Kozlov M, Markovtsov V, Kashlev M, Heumann H, Nikiforov V, Goldfarb A, Mustaev A. Mapping of catalytic residues in the RNA polymerase active center. Science. 1996;273:107–109.
14. Carpousis AJ, Gralla JD. Interaction of RNA polymerase with lacUV5 promoter DNA during mRNA initiation and elongation. Foot-printing, methylation, and rifampicin-sensitivity changes

accompanying transcription initiation. J Mol Biol. 1985;183:165-77.
15. Sydow JF, Brueckner F, Cheung AC, Damsma GE, Dengl S, Lehmann E, Vassylyev D, Cramer P: Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. Mol Cell. 2009,34:710-721.
16. Zenkin N, Yuzenkova Y, Severinov K. Transcript-Assisted Transcriptional Proofreading. Science. 2006;313:518-520.
17. Zhou M, Law JA, RNA Pol IV and V in Gene Silencing: Rebel Polymerases Evolving Away From Pol II's Rules. Curr Opin Plant Biol. 2015;27:154–164. DOI: 10.1016/j.pbi.2015.07.005
18. Kuhn CD, Geiger SR, Baumli S, Gartmann M, Gerber J, Jennebach S, Mielke T, schochner HT, Beckmann R, Cramer P. Functional Architecture of RNA Polymerase I. Cell. 2007;131:1260–1272,
19. Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD. Structural Basis of Transcription: An RNA polymerase II Elongation Complex at 3.3 Å Resolution; 2001. Acailable:www.sciencexpress.org 10.1126/science.1059495
20. Engel C, Sainsbury S, Cheung AC, Kostrewa D, Cramer P. RNA polymerase I structure and transcription regulation Nature. 2013;502:650-656.
21. Russell J, Zomerdijk JCBM. The RNA polymerase I transcription machinery. Biochem Soc Symp. 2006;73:203–216.
22. Vanittanakom N, Cooper CR, Fisher MC, Sirisanthana T *Penicillium marneffei* infection and recent advances in the epidemiology and molecular biology aspects. Clinical Microbiol Rev. 2006; 19:95–110.
23. Cramer P. Multisubunit RNA polymerases. Curr Opin Struct Biol. 2002;12:89-97.
24. Donaldson IM, Friese JD. Zinc Stoichiometry of Yeast RNA Polymerase II and Characterization of Mutations in the Zinc-binding Domain of the Largest Subunit. J Biol Chem. 2000;275:13780–13788.
25. Nudler, E., Avetissova, E., Markovtsov, V., and Goldfarb, A. Transcription processivity: protein-DNA interactions holding together the elongation complex. Science. 1996;273:211–217.
26. Geiduschek EP, Kassavetis GA. The RNA polymerase III transcription apparatus. J. Mol. Biol. 2001;310:1-26.
27. Whitehall SK, Bardeleben C, Kassavetis GA. Hydrolytic cleavage of nascent RNA in RNA polymerase III ternary transcription complexes. J Biol Chem. 1994;269:2299-306.
28. Chédin S, Riva M, Schultz P, Sentenac A, Carles C. The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. Genes Dev. 1998;12:3857–3871.
29. Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vasquez J, Hakimi MA, Lerbs-Mache S, Colot V, Lagrange T. Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. Genes Dev. 2005;19:2030–2040.
30. Marasco M, Li W, Lynch M, Pikaard CS. Catalytic properties of RNA polymerases IV and V: accuracy, nucleotide incorporation and rNTP/dNTP discrimination. Nuc Acids Res. 2017;45:11315–11326.
31. Haag JR, Pikaard CS. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nature Rev. Mol Cell Biol. 2011;12:483-492.
32. Xie M, Yu B. siRNA-directed DNA Methylation in Plants. Curr Genomics. 2015;16:23-31.
33. Kostyuk SM, Dragan DL, Lyakhov VO, Rechinsky VL, Tunitskaya BK. Chernov SN, Kochetkov E. Mutants of T7 RNA polymerase that are able to synthesize both RNA and DNA. FEBS Lett. 1995;369:165–168.
34. Palanivelu P. RNA-Dependent RNA Polymerases of Severe Acute Respiratory Syndrome-Related Coronaviruses- An Insight into their Active Sites and Mechanism of Action. Int J Biochem Res Rev. 2020;29:29-52.
35. Coleman JE. in Zinc Enzymes (Spiro TG., editor). John Wiley and Sons. Toronto; 1983.
36. Palanivelu P. An overview of the proofreading functions in bacteria and SARS-Coronaviruses. Int J Biochem Res Rev. 2021;30:33-62.
37. Sosunov V, Sosunova E, Mustaev A, Bass I, Nikiforov V, Goldfarb A. Unified two-metal mechanism of RNA synthesis and

degradation by RNA polymerase. EMBO Journal. 2003;22:2234-2244.
38. Cramer P, Bushnell DA, Kornberg RD. Structural Basis of Transcription: RNA Polymerase II at 2.8 Angstrom Resolution. Science. 2001;292:1863-1876.
39. Svetlov V, Vassylyev DG, Artsimovitch I. Discrimination against deoxyribonucleotide substrates by bacterial RNA polymerase. J Biol Chem. 2004;279:38087-38090.
40. Trinh V, Langelier MF, Archambault J, Coulombe B. Structural Perspective on Mutations Affecting the Function of Multisubunit RNA Polymerases. Microbiol Mol Biol Rev. 2006;70:12–36.
41. Thomas MJ, Platas AA, Hawley DK: Transcriptional fidelity and proofreading by RNA polymerase II. Cell. 1998;93:627-637.
42. Weilbaecher RG, Awrey DE, Edwards AM, Kane CM. Intrinsic Transcript Cleavage in Yeast RNA Polymerase II Elongation Complexes. J Biol Chem. 2003;278: 24189–24199.
43. Jeon C and Agarwal K. Fidelity of RNA polymerase II transcription controlled by elongation factor TFIIS. Proc Nat Acad Sci. (USA). 1996;93:13677-13682.
44. Gu W and Reines D. Variation in the Size of Nascent RNA Cleavage Products as a Function of Transcript Length and Elongation Competence. J Biol Chem. 1995;270:30441–30447.
45. Wang D, Bushnell DA, Huang X, Westover KD, Levitt M, and Kornberg RD. Structural basis of transcription: Backtracked RNA polymerase II at 3.4 Å resolution. Science. 2009;29(324):1203–1206.



*Peer-review history:*
*The peer review history for this paper can be accessed here:*
*http://www.sdiarticle4.com/review-history/75573*