Biotechnology Journal International

20(3): 1-35, 2017; Article no.BJI.38632
ISSN: 2456-7051
(Past name: British Biotechnology Journal, Past ISSN: 2231–2927, NLM ID: 101616695)

Single Subunit RNA Polymerases: An Insight into their Active Sites and Catalytic Mechanism

P. Palanivelu[1*]

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

Author's contribution

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

Article Information

DOI: 10.9734/BJI/2017/38632
Editor(s):
(1) Sonal S. Joshi, Department of Biological Sciences, Wayne State University, Detroit, Michigan, USA.
Reviewers:
(1) Michael G. Mauk, Drexel University, College of Engineering, USA.
(2) Yehia A. Osman Ellazeik, Mansoura University, Egypt.
(3) Roumiana Todorova, Bulgaria.
(4) Katharina Semrad, University of Vienna, Austria.
Complete Peer review History: http://www.sciencedomain.org/review-history/22801

Original Research Article

Received 4th November 2017
Accepted 15th January 2018
Published 20th January 2018

ABSTRACT

Aim: To analyze various single subunit DNA dependent RNA polymerases and identify conserved motifs, active site regions among them and propose a plausible mechanism of action for these polymerases using the T7 RNA polymerase as a model system.
Study Design: Bioinformatics, Biochemical, Site-directed mutagenesis and X-ray crystallographic data were analyzed.
Place and Duration of Study: Department of Molecular Microbiology, School of Biotechnology, Madurai Kamaraj University, Madurai – 625 021, India, from 2010 to 2013.
Methodology: The advanced version of Clustal Omega was used for protein sequence analysis of various SSU DNA dependent RNA polymerases from viruses, mitochondria and chloroplasts. Along with the conserved motifs identified by the bioinformatics analysis and with the data obtained by X-ray crystallographic, biochemical and site-directed mutagenesis were also used to confirm the possible amino acids involved in the active sites and catalysis of these RNA polymerases.
Results: Multiple sequence analyses of various single subunit (SSU) DNA dependent RNA polymerases from different sources showed only a few highly conserved motifs among them, except chloroplast RNA polymerases where a large number of highly conserved motifs were found.

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

Possible catalytic regions in all these polymerases consist of a highly conserved amino acid K and a 'gate keeper' YG pair. In addition to, these polymerases also use an invariant R at the -4 position from the YG pair and an invariant S/T, adjacent to the YG pair. Furthermore, two highly conserved Ds are implicated in the metal binding site and thus might participate in the catalytic process. The YG pair appears to be specific for DNA templates as it is not reported in RNA dependent RNA polymerases.

Conclusion: The highly conserved amino acid K, the 'gate keeper' YG pair and an invariant R which are reported in all DNA polymerases, are also found in these DNA dependent RNA polymerases. Therefore, these RNA polymerases might be using the same catalytic mechanism like DNA polymerases. The catalytic amino acid K could act as the proton abstractor and generate the necessary nucleophile at the 3'-OH and the YG pair, R and the S/T might involve in the template binding and selection of nucleoside triphosphates (NTPs) for polymerization reactions. The two highly conserved Ds could act as the ' NTP charge shielder' and orient the alpha phosphate of incoming NTPs for reaction at the 3'-OH growing end.

Keywords: Single subunit DNA dependent RNA polymerases; viral RNA polymerases; Mitochondrial RNA polymerases; chloroplast RNA polymerases; clustal omega; conserved motifs; polymerase active site; RNA polymerase mechanism.

1. INTRODUCTION

RNA polymerases (EC 2.7.7.6) are one of the key enzymes that participate in the flow of genetic information in all organisms and they play vital role in copying DNA sequences into RNA sequences, which are subsequently translated into proteins which are the final players in the cellular processes. The process of copying the DNA into RNA by RNA polymerases is known as transcription. Though RNA polymerases are found in all species, their number and composition vary across taxa. For instance, viruses contain mainly two types of RNA polymerases, viz. DNA dependent RNA polymerases and RNA dependent RNA polymerases depending upon their genetic material. Bacteria contain a single type of RNA polymerase, a multi-subunit enzyme composed of 6 subunits, while eukaryotes contain five distinct types RNA polymerases which are also multi-subunit (made up of up to 12 subunits) enzymes. In spite of such differences, there are striking similarities among transcriptional mechanisms by these polymerases.

1.1 Discovery of RNA Polymerase

RNA polymerase which makes mRNAs in the cells was discovered independently by Charles Loe, Audrey Stevens, and Jerard Hurwitz in 1960 [1]. The Nobel Prize in physiology or medicine was awarded (1959) to Severo Ochoa 'for his discovery of the mechanisms in the biological synthesis of RNA' and in chemistry was awarded (2006) to Roger D. Kornberg 'for his studies of the molecular basis of eukaryotic transcription'.

1.2 Dynamics of RNA Polymerization

RNA polymerases belong to the Main class 'Transferases' and are involved in the transfer of ribonucleoside triphosphates (NTPs/rNTPs) (Fig. 1a). Although the transcribed RNA contains the same genetic information of its DNA template, yet it is not an identical copy of the DNA segment, i.e., its sequence is only complementary to the DNA template and all the thymine residues are replaced by uracil residues in the RNA sequences (Fig. 1b) which helps the ribosomes to read the genetic code during the translation process.

RNA polymerases bind to the 3' end of a gene (promoter) and read the template DNA from 3' to 5' direction and thus a new strand is synthesized in the 5' to 3' direction. RNA polymerases direct initiation (usually initiate with a G) and catalyze further elongation at the 3'-end of an RNA by one nucleotide at a time (Figs. 1a and 1b). Unlike DNA polymerases they can initiate a chain *'de novo'* (i.e.), they do not require a primer. The overall reaction catalyzed by an RNA polymerase may be written as,

$$\text{DNA/RNA Template}_n\text{—3'OH + NTPs} \xrightarrow[Mg^{2+}]{\textit{RNA polymerase}} \text{RNA}_{n+1}\text{—3'OH + PP}_i$$

Fig. 1a. Dynamics of RNA polymerization

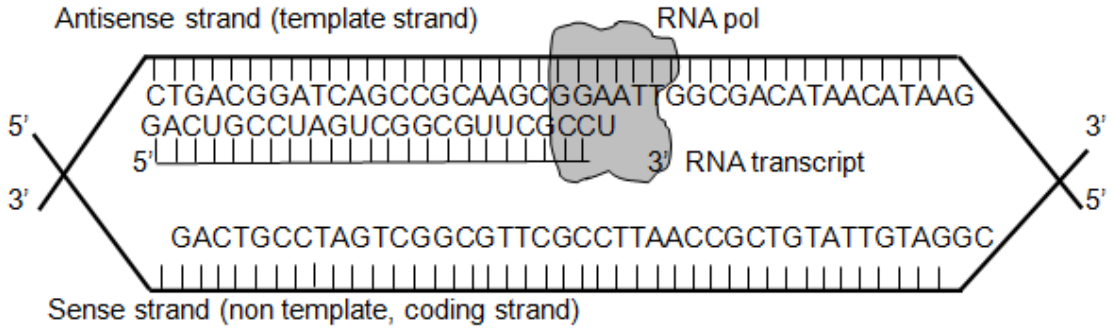


Fig. 1b. Process of RNA transcription using the template strand

Figure legends 1a: In an NTP, the strong negative charges on the phosphate groups repel each other and hence weakens the P—O bond. The hydrolysis of P—O bond results in the release of large negative free energy, which is utilized in the formation of the phosphodiester bond involving large positive free energy in DNA and RNA polymerases in general.

The RNA synthesis involves three steps, viz. initiation, elongation and termination. The newly formed RNA copies of the gene serve as blueprints for protein synthesis during the next step of translation. The basic transcription unit is the distance between the sites of Transcription Start Site (TSS) and Termination site, and may have one or more genes, e.g., mono or poly-cistronic.

As the prokaryotic and eukaryotic enzymes are multi-subunit enzymes and more complex, in this communication only the SSU (SSU) DNA dependent RNA polymerases are analyzed.

1.3 Types of SSU DNA Dependent RNA Polymerases

There are at least 3 different SSU DNA dependent RNA polymerases. They are:

1) SSUDNA dependent RNA polymerases of viruses (T7, T3, SP6, K11, etc)
2) SSU DNA dependent RNA polymerases of chloroplast
3) SSU DNA dependent RNA polymerases of mitochondria

In this communication, these three different SSU RNA polymerases are analyzed for their conserved motifs, active sites, metal binding sites and from these findings, a plausible mechanism of action is proposed for these types of enzymes.

1.3.1 Viral RNA polymerase

Depending upon the genome, the viruses are classified into two major types, viz. DNA and RNA viruses. Thus, they use DNA dependent RNA polymerases and RNA dependent RNA polymerases, respectively.

Many of these viruses use a single-subunit DNA-dependent RNA polymerase or RNA dependent RNA polymerases. The single-subunit DNA-dependent RNA polymerases especially from T7, T3, SP6 and K11 are structurally and mechanistically similar to the single-subunit RNA polymerases of eukaryotic chloroplasts and mitochondria, and are closely related to DNA polymerases (EC 2.7.7.7.) and reverse transcriptases (EC 2.7.7.49) [2].

1.3.2 Mitochondrial RNA polymerases

Mitochondria contain a single type of DNA-dependent RNA polymerases and they are single-subunit enzymes which are structurally and mechanistically very similar to the single-subunit viral DNA-dependent RNA polymerases.

1.3.3 Chloroplast RNA polymerases

Chloroplasts contain two types of DNA-dependent RNA polymerases. For example, plastids in photosynthetic higher plants use two different RNA polymerases. A multi-subunit one, very similar to bacterial RNA polymerases which is composed of α-, β-, β′-, and β"-subunits encoded by *rpoA*, *rpoB*, *rpoC1,* and *rpoC2* genes, and is referred to as plastid-encoded polymerase (PEP). The second enzyme is referred to as the nucleus-encoded polymerase (NEP). The NEPs is structurally and mechanistically unrelated to PEP but belong to the "single-subunit RNA polymerases" protein family. Interestingly, the NEPs are very similar to

the viral RNA polymerases of T3, T7, SP6, K11, etc.

2. MATERIALS AND METHODS

A large number of RNA polymerases from various organisms have been isolated, purified, characterized, cloned and sequenced [3 and references therein]. Complete nucleic acid and protein sequence data are available for many of these enzymes from different species. Thus, these data have become valuable tools in analyzing and understanding the structure-function relationships of these enzymes. This communication presents the results obtained from the protein sequence analysis of these enzymes, which are supported by biochemical, site-directed mutagenesis experiments and X-ray studies data on these enzymes.

The bacteriophage T7 DNA-dependent RNA polymerase is used as the model system for delineating the polymerization mechanism. Particular features of this enzyme, viz. the single-subunit composition, relatively low molecular weight and large amount of data on biochemical, site-directed mutagenesis and X-ray analyses make this enzyme the most convenient model for investigating the physicochemical aspects of transcription. For multiple sequence analysis (MSA) of various RNA polymerases, the sequences were retrieved from SWISS-PROT and PUBMED sites and analyzed using Clustal Omega, an accurate, fast and widely accepted algorithm, available on their website.

3. RESULTS AND DISCUSSION

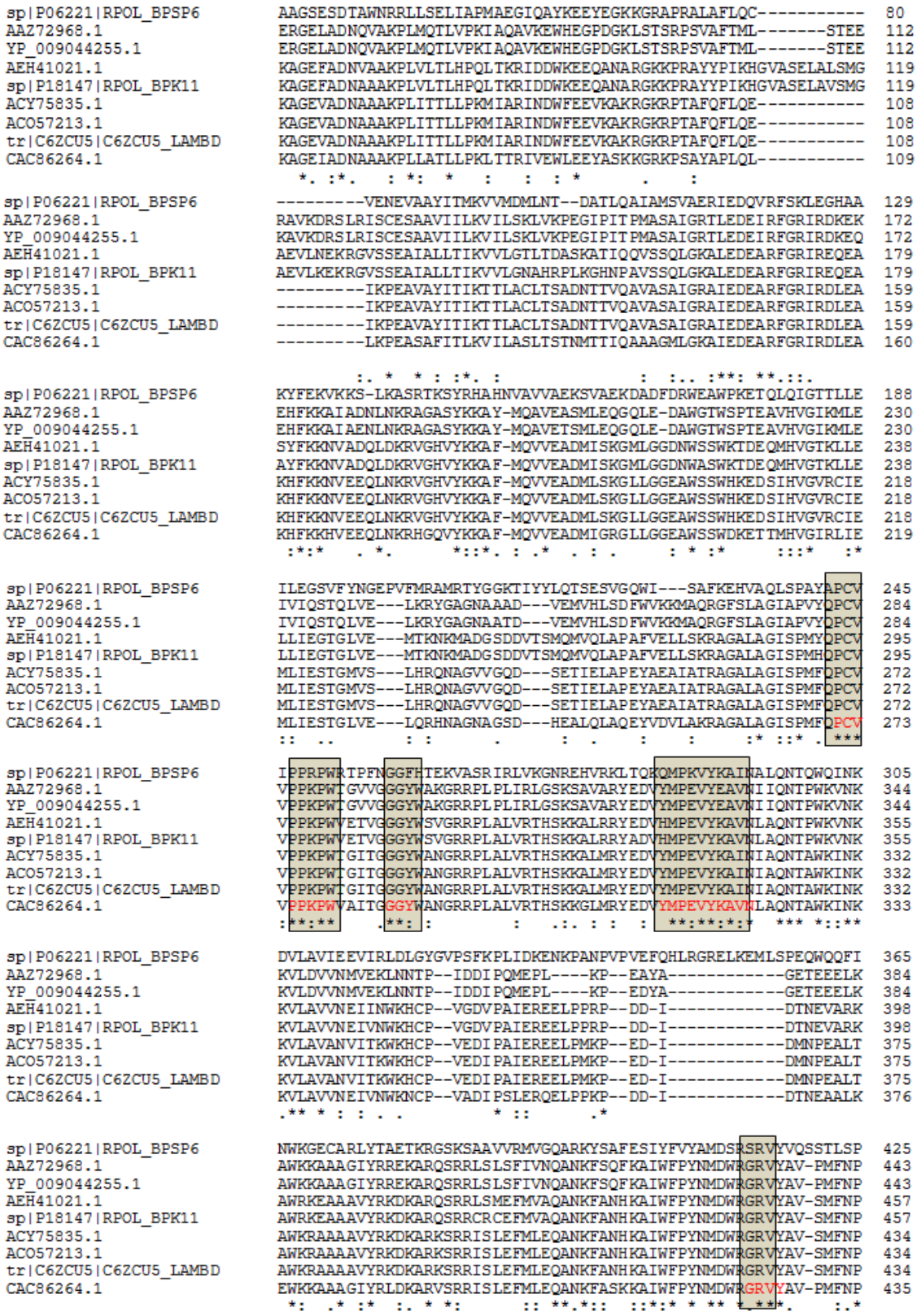
Figs. 2-7 show the MSA of various polymerases and their combinations (only the relevant and highly conserved regions are shown).

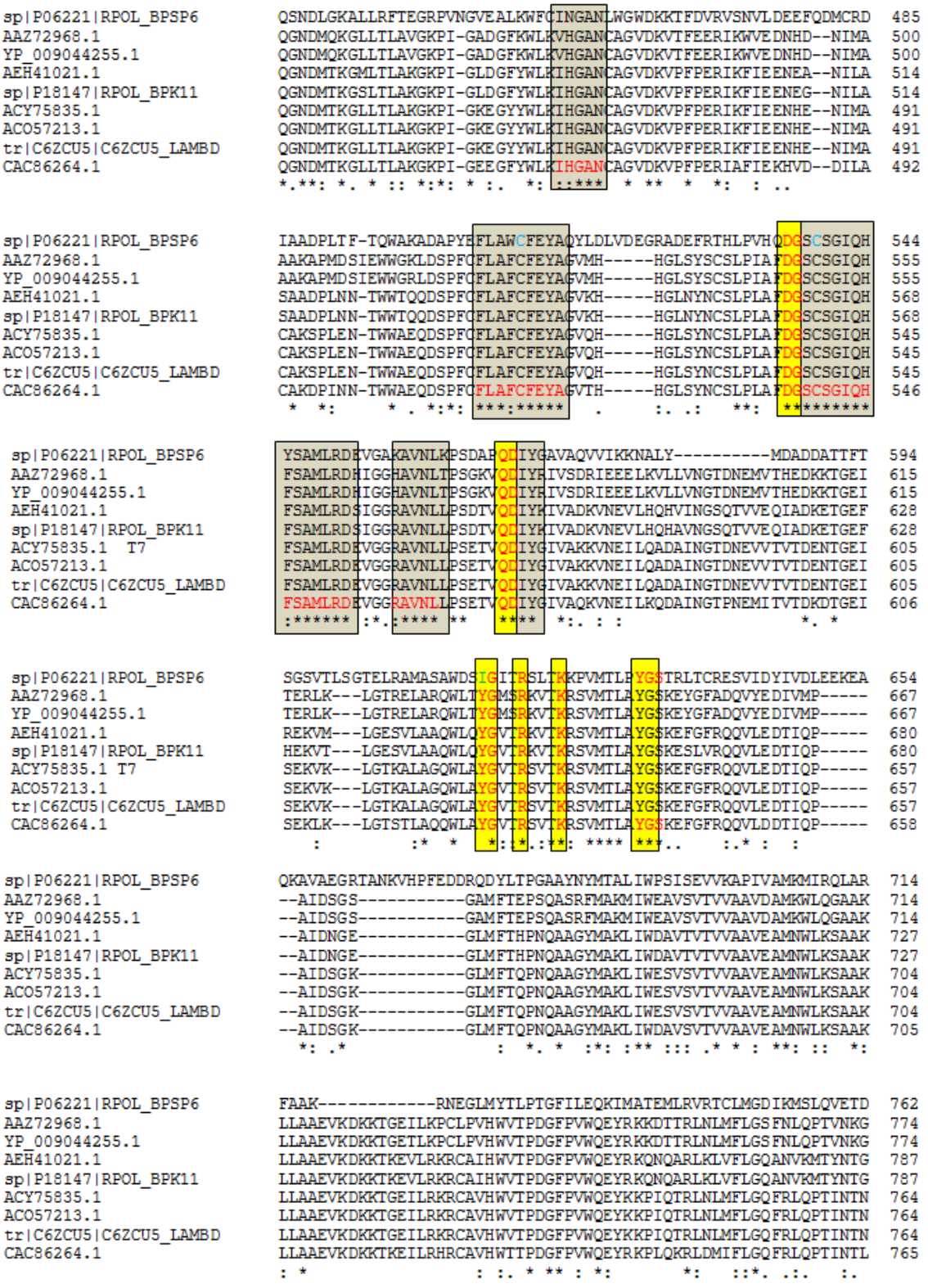
3.1 MSA of SSU RNA Polymerases from Different Sources

Fig. 2 shows the multiple sequence alignment and conserved motifs in SSU viral RNA polymerases such as T3, T7, SP6, K11, etc. There are large numbers of conserved motifs and amino acids among them. The catalytic, template and substrate binding motifs are highlighted. The YG 'gate keeper' motif and the catalytic K are strictly conserved (including distance conservation) DNA dependent RNA polymerases from the viruses. Similar observations were made in DNA dependent DNA polymerases also [4]. This strongly suggests that the DNA and RNA polymerases use the same set of amino acids for template, substrate binding and catalysis establishing a structure-function relationship among the DNA and RNA polymerases. The immediate downstream amino acid in DNA polymerases is usually a G or A [4] but in viral RNA polymerases it is a K or R. Interestingly, an R is found (4th amino acid downstream from the catalytic K) as the invariant amino acid in both the DNA and RNA polymerases. Another interesting observation is that the RNA polymerases from the enterobacteriophages possess one more YG 'gate keeper' pair exactly at the same distance but from the downstream of the catalytic K (The SP6 polymerase slightly deviates from others and uses TG). This suggests that the two YG pairs might be recognizing and binding on both the coding and non-coding templates of the DNA (please note that the DNA polymerases use only one template and shows one YG pair and uses a primer) whereas the catalytic K positioning in the middle might be catalyzing the NTP addition. In fact, these RNA polymerases require a double-stranded DNA for transcription as no activity was found when the T7 RNA polymerase was assayed on single-stranded DNA as substrate [5].

CLUSTAL O (1.2.4) MSA of DNA dependent RNA polymerases from enterobacteriophages like T3, T7, K11, SP6 ,λ, etc

```
sp|P06221|RPOL_BPSP6        ----------------------------MQDLHAIQLQLEEEMFNGGIRRFEADQQRQI   31
AAZ72968.1                  -MSVISIDKHDFSDVSNAIEPFNLLADHYGQDLAVKQLQLEHEAYTEGERRFIKNLERQT   59
YP_009044255.1              -MSVISIDKHDFSDVSNAIEPFNLLADHYGQDLAVKQLQLEHEAYTEGERRFIKNLERQT   59
AEH41021.1                  -MNALNIARNDFSEIELAAIPYNILSEHYGDRLAREQLALEHEAYELGEQRFLKMLERQV   59
sp|P18147|RPOL_BPK11        -MNALNIGRNDFSEIELAAIPYNILSEHYGDQAAREQLALEHEAYELGRQRFLKMLERQV   59
ACY75835.1                  -MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRKMFERQL   59
ACO57213.1                  -MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRKMFERQL   59
tr|C6ZCU5|C6ZCU5_LAMBD      -MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRKMFERQL   59
CAC86264.1                  MNIIENIEKNDFSEIELAAIPFNTLADHYGSALAKEQLALEHESYELGERRFLKMLERQA   60
                                                  .     ** **.* :  *  **    :**
```


sp|P06221|RPOL_BPSP6
AAZ72968.1
YP_009044255.1
AEH41021.1
sp|P18147|RPOL_BPK11
ACY75835.1
ACO57213.1
tr|C6ZCU5|C6ZCU5_LAMBD
CAC86264.1


sp|P06221|RPOL_BPSP6
AAZ72968.1
YP_009044255.1
AEH41021.1
sp|P18147|RPOL_BPK11
ACY75835.1
ACO57213.1
tr|C6ZCU5|C6ZCU5_LAMBD
CAC86264.1
ACY75835.1 T7

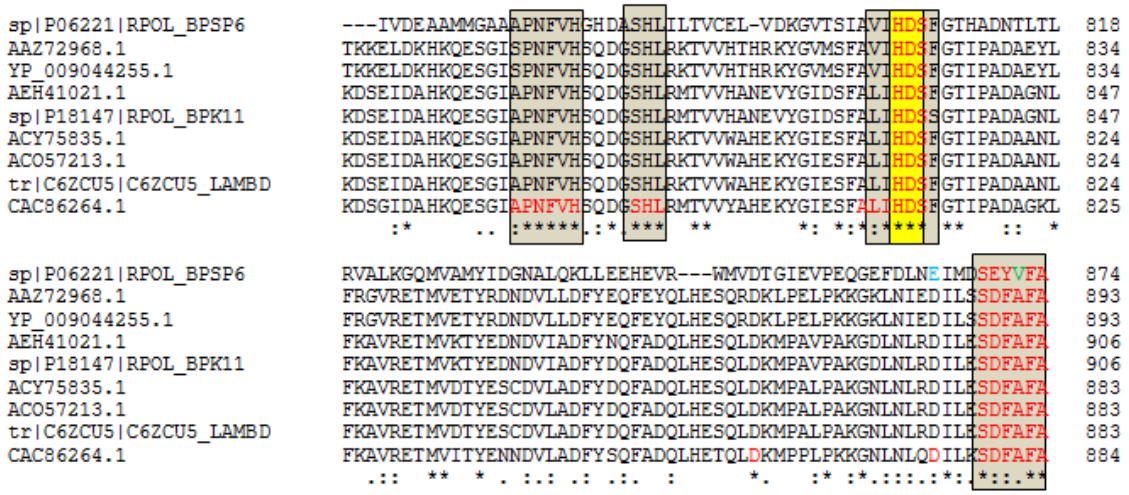
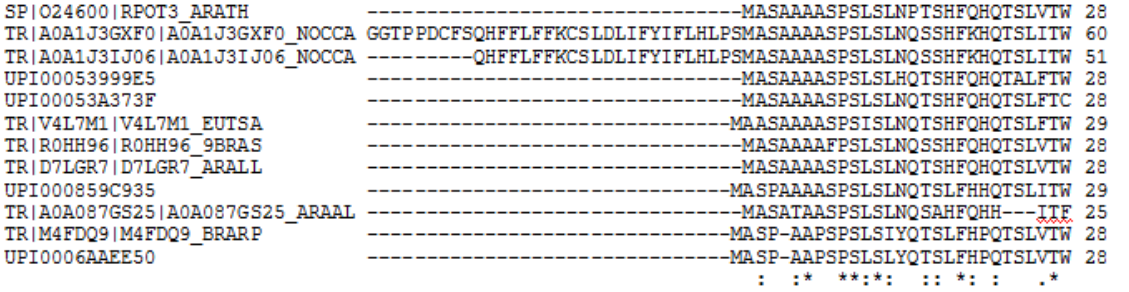


Fig. 2. MSA of SSU RNA polymerases from enterobacteria phages like T3, T7, SP6, K11, λ, etc

sp|P06221|RPOL_BPSP6 DNA-directed RNA polymerase OS=Enterobacteria phage SP6
AAZ72968.1 RNA polymerase [Enterobacteria phage K1F]
YP_009044255.1 RNA polymerase [Escherichia phage PE3-1]
AEH41021.1 RNA polymerase [Escherichia phage K30]
sp|P18147|RPOL_BPK11 DNA-directed RNA polymerase OS=Enterobacteria phage K11
ACY75835.1 T7 RNA polymerase [Enterobacteria phage T7]
ACO57213.1 RNA polymerase [Enterobacteria phage T7]
tr|C6ZCU5|C6ZCU5_LAMBDA DNA-directed RNA polymerase OS=Escherichia phage
CAC86264.1 RNA polymerase [Enterobacteria phage T3]

Fig. 3 shows the MSA and conserved motifs in SSU chloroplast RNA polymerases. As compared to mitochondrial RNA polymerases, there are large regions of conserved regions among chloroplast RNA polymerases; they are about 90% homologous. The catalytic, template and substrate binding motifs are highlighted. The YG gate keeper motif and the catalytic K is strictly conserved (including distance conservation) in SSU chloroplast RNA polymerases also. This strongly suggests that the DNA and RNA polymerases might be using the same set of amino acids for template, substrate binding and catalysis. The immediate downstream amino acid in DNA polymerases is usually a G or A [4] but in these RNA polymerases, it is a Q. Here also an R is found as the invariant amino acid at -4 position from the catalytic K. Another interesting feature in these RNA polymerases is that a YG pair is present like the viral RNA polymerases but much upstream from the catalytic K.

CLUSTAL O(1.2.4) MSA of SSU RNA polymerases from chloroplasts



```
SP|O24600|RPOT3_ARATH          LKPPPSS--ALFRRKTLPFFERHSLPIS--ASS-SSS--SSSSTSLSVHEKPISNSVHFH 81
TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA LKPPSSS--ALFRRKILPLSSPVR---RLSLPISASSSS-SSSTSLSVSEKPTANSVHFH 114
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA LKPPSSS--ALFRRKILPLSSPVR---RLSLPISASSSSSSSSTSLSVSEKPTANSVHFH 106
UPI00053999E5                  LKPPSSY--ALFRRKTLPFFGRHSLPIS--ASS-SSSSSSSSTSLSVHEKPISNSVHFH 83
UPI00053A373F                  LKPPSSP--ALFRRKTLPFFQRHSLPIS--ASS-SSSSSSSSTSLSVHEKPISNSVHFH 83
TR|V4L7M1|V4L7M1_EUTSA         LKPPSS---ALFRRKTLPFFDPSSPDKRISLPP-ISASSSSSSASLSVSEKP---TVHFH 82
TR|R0HH96|R0HH96_9BRAS         LKPP-SS--ALYRRKTLPFFERHSLPIS--ASS-SS---SSSSTSLSVHEKPISNSVHFH 79
TR|D7LGR7|D7LGR7_ARALL         LKPPPSS--ALFRRKTLPFFERHSLPIS--ASS-SS-----SSSSLSVHEKPISNSVHFH 78
UPI000859C935                  LKPPSSSSSSALFRRKATKRLL---------PPI-SAASSS-SSTSLSV--TTEKPTVHFH 76
TR|A0A087GS25|A0A087GS25_ARAAL LKPPTSS--TLLRRKTLPFSVKR--------ISASSSSSSSSTSLSVTEKPTSNSVHFH 75
TR|M4FDQ9|M4FDQ9_BRARP         LKPPSS--SALFRRKPTKRLL---------PPI-SAASSSSSSTSLSL--TEKPTTVHFH 74
UPI0006AAEE50                  LKPPSS--SALFRRKPTKRLL---------PPI-SAASSSSSSTSLSV--TEKPTTVHFH 74
                               **** *   :* ***                  :     **:***:  .    :****

SP|O24600|RPOT3_ARATH          GNLIESFENQDSSYAGTIKGASLIEELENPVERNGLSGRRRLFMQDPPWISALFLKGLSK 141
TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA GNLIESFENQD-SFAGSINGTSLIDELENPVQRNGISGRRRLFMQDPPWISALFLKGLSK 173
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA GNLIESFENQD-SFAGSINGTSLIDELENPVQRNGISGRRRLFMQDPPWISALFLKGLSK 165
UPI00053999E5                  GNLIESFENQDSSYAGTIKGASLVEELENPVERNGLSGRRRLFMQDPPWISALFLKGLSK 143
UPI00053A373F                  GNLIESFENQDSSYAGAIKGASLVEELENPVERNGLSGRRRLFMQDPPWISALFLKGLSK 143
TR|V4L7M1|V4L7M1_EUTSA         GNLIESFENQD-SFAGTINGTSLIEELENPVERNGLSGRRRLFMQDPPWISALFLKGLSK 141
TR|R0HH96|R0HH96_9BRAS         GNLIESFENQDSSYAGTIKGASLIEELENPVERNGLSGRRRLFMQDPPWISALFLKGLSK 139
TR|D7LGR7|D7LGR7_ARALL         GNLIESFENQDSSYAGTINGASLIEELENPVERNGLSGRRRLFMQDPPWISALFLKGLSK 138
UPI000859C935                  GNLIESFESH----AGTIKGATF---TDNPVERNELSARKRLFTQDPPWISALFLKGLTK 129
TR|A0A087GS25|A0A087GS25_ARAAL GNLIDSFENQSG----TIKGATL---IENPVERSELSGRRRLFMQDPPWISALFLKGLTK 128
TR|M4FDQ9|M4FDQ9_BRARP         GNLIESFESHD-SFPGAIKGAAF---TENPVERSELSATRRLFTQDPPWISALFLKGLTK 130
UPI0006AAEE50                  GNLIESFEGHD-SFPGAIKGAAF---TENPVERSELSATRRLFTQDPPWISALFLKGLTK 130
                               ****:***.:     :*:*:::    :***:*. :*. :*** **************:*

SP|O24600|RPOT3_ARATH          MV-DQTLKIERKDIDKRKFDSLRRRQVKEETEAWERMVDEYRDLEKEMCEKNLAPNLPYV 200
TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA MV-DQTLKIERKDIDKRKFDSLRRRQVKEETEAWERMVDEYRDLEKEMCEKNLAPNLPYV 232
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA MV-DQTLKIERKDIDKRKFDSLRRRQVKEETEAWERMVDEYRDLEKEMCEKNLAPNLPYV 224
UPI00053999E5                  MV-DQTVKIEHKDIDKRKFDSLRRRQVKEETEAWEKMVDEYRDLEKEMCEKNLAPNLPYV 202
UPI00053A373F                  MV-DQTVKIEHKDIDKRKFDSLRRRQVKEETEAWEKMVDEYRDLEKEMCEKNLAPNLPYV 202
TR|V4L7M1|V4L7M1_EUTSA         MV-DQTVKIERKDIDKRKFDSLRRRQVKEETEAWEKMVDEYRDLEKEMCEKNLAPNLPYV 200
TR|R0HH96|R0HH96_9BRAS         MV-DQTVKIEHKYIDKRKFDSLRRRQVKEETEAWEKMVDEYRDLEKEMCEKNLAPNLPYV 198
TR|D7LGR7|D7LGR7_ARALL         MV-DQTLKIERKDIDKRKFDSLRRRQVKEETEAWERMVDEYRDLEKEMCEKNLAPNLPYV 197
UPI000859C935                  MAADQTVKIERKDIDKRKFDSLRRRQVKEETEAWERTVDEYRDLEKEMCEKSLAPSLPYV 189
TR|A0A087GS25|A0A087GS25_ARAAL MV-DQTFQIERKDIDKRKFDLLRRRQVKEETEAWEKMVDEYRDLEKEMCEKNLAPNLPYV 187
TR|M4FDQ9|M4FDQ9_BRARP         M----TVKIERKDIDKRKFDSLRRRQVKEETEAWERMVEEYRDLEKEMCEKSLAPNLPYV 186
UPI0006AAEE50                  M----TVKIERKDIDKRKFDSLRRRQVKEETEAWERMVEEYRDLEKEMCEKSLAPNLPYV 186
                               *    *.:**:* ******* **************: *:************.***.****

SP|O24600|RPOT3_ARATH          KHMFLGWFQPLKDVIEREQKLQKNKSKKVRAAYAPHIELLPADKMAVIVMHKMMGLVMSG 260
TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA KHMFLGWFQPLKDVIEREQKLQKNKSKKVRAAYAPHIELLPADKMAVIVMHKMMGLVMSG 292
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA KHMFLGWFQPLKDVIEREQKLQKNKSKKVRAAYAPHIELLPADKMAVIVMHKMMGLVMSG 284
UPI00053999E5                  KHMFLGWFQPLKDVIEREQKLQKNKSKKVRAAYAPHIELLPADKMAVIVMHKMMGLVMSG 262
UPI00053A373F                  KHMFLGWFQPLKDVIEREQKLQKNKSKKVRAAYAPHIELLPADKMAVIVMHKMMGLVMSG 262
TR|V4L7M1|V4L7M1_EUTSA         KHMFLGWFQPLKDVIEREQKLQKNKSKKVRAAYAPHIELLPADKMAVIVMHKMMGLVMSG 260
TR|R0HH96|R0HH96_9BRAS         KHMFLGWFQPLKDVIEREQKLQKNKSKKVRAAYAPHIELLPADKMAVIVMHKMMGLVMSG 258
TR|D7LGR7|D7LGR7_ARALL         KHMFLGWFQPLKDVIEREQKLQKNKSKKVRAAYAPHIELLPADKMAVIVMHKMMGLVMSG 257
UPI000859C935                  KHMFLGWFQPLKEVIEREQKLQKNKSKKVRAAYAPHIELLPADKMAVIVMHKMMGLVMSG 249
TR|A0A087GS25|A0A087GS25_ARAAL KHMFLGWFQPLKDVIEREQKLQKNKSKKVRAAYAPHIELLPADKMAVIVMHKMMGLVMSG 247
TR|M4FDQ9|M4FDQ9_BRARP         KHMFLGWFQPLKEVIEREQRLQKNKSKKVRAAYAPHIELLPADKMAVIVMHKMMGLVMSG 246
UPI0006AAEE50                  KHMFLGWFQPLKEVIEREQRLQKNKSKKVRAAYAPHIELLPADKMAVIVMHKMMGLVMSG 246
                               ************:******:****************************************

SP|O24600|RPOT3_ARATH          HEDGCIQVVQAAVSIGIAIEQEVRIHNFLKRTRKNNAGDSQEELKEKQLLRKRVNSLIRR 320
TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA HEDGCIQVVQAAVSIGVAIEQEVRIHNFLKRTRKNNEGDSPEELKDKQLLRKRVNSLIRR 352
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA HEDGCIQVVQAAVSIGVAIEQEVRIHNFLKRTRKNNEGDSPEELKDKQLLRKRVNSLIRR 344
UPI00053999E5                  HEDGCIQVVQAAVSIGIAIEQEVRIHNFLKKTRKNNAGDSQEELKEKQLLRKRVNSLIRR 322
UPI00053A373F                  HEDGCIQVVQAAVSIGIAIEQEVRIHNFLKRTRKNNAGDSQEELKEKQLLRKRVNSLIRR 322
TR|V4L7M1|V4L7M1_EUTSA         HEDGCIQVVQAAVSIGIAIEQEVRIHNFLKRTRKNNAGDSEEELKDKQLLRKRVNSLIRR 320
TR|R0HH96|R0HH96_9BRAS         HEDGCIQVVQAAVSIGIAIEQEVRIHNFLKRTRKNNAGDSQEELKEKQLLRKRVNSLIRR 318
TR|D7LGR7|D7LGR7_ARALL         HEDGCIQVVQAAVSIGIAIEQEVRIHNFLKRTRKNNAGDSQEELKEKQLLRKRVNSLIRR 317
UPI000859C935                  HEDGCIQVVQAAVSIGIAIEHEVRIHNFLKRSRKNNAGDSEEELKDKQLLRKRVNSLIRR 309
TR|A0A087GS25|A0A087GS25_ARAAL HEDGCIQVVQAAVSIGIAIEHEVRIHNFLKRTRKNNGGDTEEELKDKQLLRKRVNGLIRR 307
TR|M4FDQ9|M4FDQ9_BRARP         HEDGCIQVVQAAVSIGIAIEHEVRIHNFLKRTRKNNAGDSQEELKDKQLLRKRVNSLIRR 306
UPI0006AAEE50                  HEDGCIQVVQAAVSIGIAIEHEVRIHNFLKRTRKNNAGDSQEELKDKQLLRKRVNSLIRR 306
                               ****************:***:*********::**** **: ****:*********.****
```

```
SP|O24600|RPOT3_ARATH           KRIIDALKVVKSEGTKPWGRATQAKLGSRLLELLIEAAYVQPPLTQSGDSIPEFRPAFRH 380
TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA KRMIDALKVVKCEGIKPWGRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 412
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA KRMIDALKVVKCEGIKPWGRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 404
UPI00053999E5                   KRIIDALKVVKSEGTKPWGRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 382
UPI00053A373F                   KRIIDALKVVKSEGTKPWGRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 382
TR|V4L7M1|V4L7M1_EUTSA          KRIIDALKVVKSEGIKPWGRATQAKLGSRLLELLIETAYVQPPLTQAGDTIPEFRPAFRH 380
TR|R0HH96|R0HH96_9BRAS          KRIIDALKVVKSEGIKPWGRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 378
TR|D7LGR7|D7LGR7_ARALL          KRIIDALKVVKTEGTKPWGRATQAKLGSRLLELLIEAAYVQPPLTQSGDSIPEFRPAFRH 377
UPI000859C935                   KRIIDALKVVKSEGIKPWGRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 369
TR|A0A087GS25|A0A087GS25_ARAAL KRIIDALKVVKCEGTKPWGRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 367
TR|M4FDQ9|M4FDQ9_BRARP          KRIIDALKVVKSEGIKPWGRATQAKLGSRLLELLMETAYVQPPLTQSGDSIPEFRPAFRH 366
UPI0006AAEE50                   KRIIDALKVVKSEGIKPWGRATQAKLGSRLLELLMETAYVQPPLTQSGDSIPEFRPAFRH 366
                                **:******** ** ******************:*:*********:**:**********

SP|O24600|RPOT3_ARATH           RFKTVTKYPGSKLVRRYGVIECDSLLLAGLDKSAKHMLIPYVPMLVPPKRWKGYDKGGYL 440
TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA KFKTVTKYPGSKLVRRYGVIECDSLLLAGLDKSAKHMLIPYVPMLVPPKRWKGYDKGGYL 472
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA KFKTVTKYPGSKLVRRYGVIECDSLLLAGLDKSAKHMLIPYVPMLVPPKRWKGYDKGGYL 464
UPI00053999E5                   TFKTVTKYPGSKLVRRYGVIECDSLLLAGLDKSAKHMLIPYVPMLVPPKRWKGYDKGGYL 442
UPI00053A373F                   TFKTVTKYPGSKLVRRYGVIECDSLLLAGLDKSAKHMLIPYVPMLVPPKRWKGYDKGGYL 442
TR|V4L7M1|V4L7M1_EUTSA          KFKTVTKYPGSKLVRRYGVIECDSLLLAGLDKSAKHMLIPYVPMLVPPKRWKGYDKGGYL 440
TR|R0HH96|R0HH96_9BRAS          KFKTVTKYPGSKLVRRYGVIECDSLLLAGLDKSAKHMIIPYVPMLVPPKRWKGYDKGGYL 438
TR|D7LGR7|D7LGR7_ARALL          KFKTVTKYPGSKMVRRYGVIECDSLLLAGLDKSAKHMLIPYVPMLVPPKRWKGYDKGGYL 437
UPI000859C935                   KFKTVTKYPGSKLVRRYGVIECDSLLLAGLDKSAKHMLIPYVPMLVPPKRWKGYDKGGYL 429
TR|A0A087GS25|A0A087GS25_ARAAL KFKTVTKYPGSKLVRRYGVIECDSLLLAGLDKSAKHMLIPYVPMLVPPKRWKGYDKGGYL 427
TR|M4FDQ9|M4FDQ9_BRARP          KFKTVTKYPGSKLVRRYGVIECDSLLLAGLDKSAKHMLIPYVPMLVPPKRWRGYDKGGYL 426
UPI0006AAEE50                   KFKTVTKYPGSKLVRRYGVIECDSLLLAGLDKSAKHMLIPYVPMLVPPKRWRGYDKGGYL 426
                                 **********:**********************:*************:********

SP|O24600|RPOT3_ARATH           FLPSYIMRTHGSKKQQDALKDISHKTAHRVFEALDTLGNTKWRVNRNILDVVERLWADGG 500
TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA FLPSYIMRTHGSKKQQDALKDISSKTAHRVFEALDTLGNTKWRVNRKILDVVEKLWADGG 532
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA FLPSYIMRTHGSKKQQDALRDISSKTAHRVFEALDTLGNTKWRVNRKILDVVEKLWADGG 524
UPI00053999E5                   FLPSYIMRTHGSKKQQDALRDISYKTAHRVFEALDTLGNTKWRVNRKILDVVERLWADGG 502
UPI00053A373F                   FLPSYIMRTHGSKKQQDALRDISHKTAHRVFEALDTLGNTKWRVNRKILDVVERLWADGG 502
TR|V4L7M1|V4L7M1_EUTSA          FLPSYIMRTHGSKKQQDALRDISSKTAHRVFEALDTLGNTKWRVNRKILDVVERLWADGG 500
TR|R0HH96|R0HH96_9BRAS          FLPSYIMRTHGSKKQQDALRDISSKTAHRVFEALDTLGNTKWRVNRKILDVVERLWADGG 498
TR|D7LGR7|D7LGR7_ARALL          FLPSYIMRTHGSKKQQDALRDISHKTAHRVFEALDTLGNTKWRVNRNILDVVERLWADGG 497
UPI000859C935                   FLPSYIMRTHGSKKQQDALKDISSKTAHRVFEALDTLGNTKWRVNRKILDVVERLWADGG 489
TR|A0A087GS25|A0A087GS25_ARAAL FLPSYIMRTHGSKKQQDALRDVSHKTAHRVFEALDTLGNTKWRVNRNILDVVERLWADGG 487
TR|M4FDQ9|M4FDQ9_BRARP          FLPSYIMRTHGSKKQQDALKDISSKTAHRVFEALDTLGNTKWRVNRKILDVVERLWADGG 486
UPI0006AAEE50                   FLPSYIMRTHGSKKQQDALKDISSKTAHRVFEALDTLGNTKWRVNRKILDVVERLWADGG 486
                                ******************:*:* **********************:******:******

SP|O24600|RPOT3_ARATH           NIAGLVNREDVPIPEKPSSEDPEELQSWKWSARKANKINRERHSLRCDVELKLSVARKMK 560
TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA NIAGLVNREDVPIPEKPTSEDPEEMQSWKWSARKANKINRERHSLRCDVELKLSVARKMK 592
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA NIAGLVNREDVPIPEKPTSEDPEEMQSWKWSARKANKINRERHSLRCDVELKLSVARKMK 584
UPI00053999E5                   NIAGLVNREDVPIPEKPSSEDPDEIQAWKWSVRKANKINRERHSLRCDVELKLSVARKMK 562
UPI00053A373F                   NIAGLVNREDVPIPEKPSSEDPDEIQAWKWSVRKANKINRERHSLRCDVELKLSVARKMK 562
TR|V4L7M1|V4L7M1_EUTSA          NIAGLVNREDVPIPEKPSSEDPEEIQSWKWSVRKANKINRERHSLRCDVELKLSVARKMK 560
TR|R0HH96|R0HH96_9BRAS          NIAGLVNREDVPIPEKPSSEDPEELQSWKWSVRKANKINRERHSLRCDVELKLSVARKMK 558
TR|D7LGR7|D7LGR7_ARALL          NIAGLVNREDVPIPEKPSSEDPEEIQSWKWSVRKANKINRERHSLRCDVELKLSVARKMK 557
UPI000859C935                   NIAGLVNREDVPIPEKPSSEDPEEIQSWKWSVRKAKKTNRERHSLRCDVELKLSVARKMK 549
TR|A0A087GS25|A0A087GS25_ARAAL NIAGLVNREDVPIPEKPSSEDPEEIQSWKWSARKANKINRERHSLRCDVELKLSVARKMK 547
TR|M4FDQ9|M4FDQ9_BRARP          NIAGLVNREDVPIPEKPSSEDPEELQSWKWSVRKAKKTNRERHSLRCDVELKLSVARKMK 546
UPI0006AAEE50                   NIAGLVNREDVPIPEKPSSEDPEELQSWKWSVRKAKKTNRERHSLRCDVELKLSVARKMK 546
                                ****************:****:*:*:****.***:* **********************

SP|O24600|RPOT3_ARATH           DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLCRGTLEFAEGRPLGKSGLHWLKIHLANLY 620
TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLCRGTLEFAEGRPLGKQGLYWLKIHLANLF 652
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLCRGTLEFAEGRPLGKQGLYWLKIHLANLF 644
UPI00053999E5                   DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLCRGTLEFAEGRPLGKSGLHWLKIHLANLY 622
UPI00053A373F                   DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLCRGTLEFAEGRPLGKSGLHWLKIHLANLY 622
TR|V4L7M1|V4L7M1_EUTSA          DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLCRGTLEFAEGRPLGKSGLYWLKIHLANLY 620
TR|R0HH96|R0HH96_9BRAS          DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLCRGTLEFAEGRPLGKSGLHWLKIHLANLY 618
TR|D7LGR7|D7LGR7_ARALL          DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLCRGTLEFAEGRPLGKSGLHWLKIHLANLF 617
UPI000859C935                   DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLCRGTLEFAEGRPLGKSGLYWLKIHLANLY 609
TR|A0A087GS25|A0A087GS25_ARAAL DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLCRGTLEFAEGRPLGKSGLHWLKIHLANLY 607
TR|M4FDQ9|M4FDQ9_BRARP          DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLCRGTLEFAEGRPLGKSGLYWLKIHLANLY 606
UPI0006AAEE50                   DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLCRGTLEFAEGRPLGKSGLYWLKIHLANLY 606
                                ********************************************.**:*********:
```

```
SP|O24600|RPOT3_ARATH           AGGVEKLSHDARLAFVENHLDDIMDSAENPIHGKRWWLKAEDPFQCLAACVILTQALKSP 680
TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA AGGVEKLSHDGRLAFVENHLDDIMESAENPVHGKRWWLKAEDPFQCLAACVILTQALKSP 712
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA AGGVEKLSHDGRLAFVENHLDDIMESAENPVHGKRWWLKAEDPFQCLAACVILTQALKSP 704
UPI00053999E5                   AGGVEKLSHDARLAFVENHLDDIMDSAENTIHGKRWWLKAEDPFQCLAACVILTQALKSP 682
UPI00053A373F                   AGGVEKLSHDARLAFVENHLDDIIDSAENTIHGKRWWLKAEDPFQCLAACVILTQALKSP 682
TR|V4L7M1|V4L7M1_EUTSA          AGGVEKLSHDGRLAFVENHLDNIMDSAENPIHGQRWWLKAEDPFQCLAACVILTQALKSP 680
TR|R0HH96|R0HH96_9BRAS          AGGVEKLSHDARLAFVETHLDDVMDSAENPIHGKRWWLKAEDPFQCLAACVILTQALKSP 678
TR|D7LGR7|D7LGR7_ARALL          AGGVEKLSHDARLAFVENHLDDIMDSAENPIHGKRWWLKAEDPFQCLAACVILTQALRSP 677
UPI000859C935                   AGGVEKLSHDGRLAFVENHLDDIIDSAENAIHGKRWWLKAEDPFQCLAACVILGQALKSP 669
TR|A0A087GS25|A0A087GS25_ARAAL AGGVEKLSHDGRLAFVENHLDVIIDSAENPIHGKRWWLKAEDPFQCLAACVILTQALKSP 667
TR|M4FDQ9|M4FDQ9_BRARP          AGGVEKLSHEGRLSFVENHLDDIMDSAENAIHGRRWWLKAEDPFQCLAACVVLAQALKSP 666
UPI0006AAEE50                   AGGVEKLSHEGRLSFVENHLDDIMDSAENAIHGRRWWLKAEDPFQCLAACVVLAQALKSP 666
                                *********:.**:***.*** :::**** :**:**************** * ***:**

SP|O24600|RPOT3_ARATH           SPYSVISHLPIHQDGSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISRRVHEIMK 740
TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA SPSSVISHLPIHQDGSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISLRVHEIMK 772
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA SPSSVISHLPIHQDGSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISLRVHEIMK 764
UPI00053999E5                   SPYSVISHLPIHQDGSCNGLQHYAALGRDSFEAAAVNLVAGDKPADVYSEISRRVHEIMK 742
UPI00053A373F                   SPYSVISHLPIHQDGSCNGLQHYAALGRDSFEAAAVNLVAGDKPADVYSEISRRVHEIMK 742
TR|V4L7M1|V4L7M1_EUTSA          SPYSVISHLPIHQDGSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISLRVHEIMK 740
TR|R0HH96|R0HH96_9BRAS          SPYSVISHLPIHQDGSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISRRVHEIMK 738
TR|D7LGR7|D7LGR7_ARALL          SPYSVISHLPIHQDGSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISRRVHEIMK 737
UPI000859C935                   SPYSVISHLPIHQDGSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISLRVHEIMK 729
TR|A0A087GS25|A0A087GS25_ARAAL SPYSVISHLPIHQDGSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISLRVHEIMK 727
TR|M4FDQ9|M4FDQ9_BRARP          SPYSVISHLPIHQDGSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISLRVHEIMK 726
UPI0006AAEE50                   SPYSVISHLPIHQDGSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISLRVHEIMK 726
                                ** ********:*****************************:********** *******

SP|O24600|RPOT3_ARATH           KDSSKDPESNPTAALAKILITQVDRKLVKQTVMTSVYGVTYVGAREQIKRRLEEKGVITD 800
TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA KDSNKDPESNPTAALAKILINQVDRKLVKQTVMTSVYGVTYVGAREQIKRRLEEKGVITD 832
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA KDSNKDPESNPTAALAKILINQVDRKLVKQTVMTSVYGVTYVGAREQIKRRLEEKGVITD 824
UPI00053999E5                   KDSSKDPESNPTAALAKILITQVDRKLVKQTVMTSVYGVTYVGAREQIKRRLEEKGVITD 802
UPI00053A373F                   KDSSKDPESNPTAALAKILITQVDRKLVKQTVMTSVYGVTYVGAREQIKRRLEEKGVITD 802
TR|V4L7M1|V4L7M1_EUTSA          KDSSKDPESNPTAALAKILINQVDRKLVKQTVMTSVYGVTYVGAREQIKRRLEEKGVISD 800
TR|R0HH96|R0HH96_9BRAS          KDSSKDPESNPTAALAKILITQVDRKLVKQTVMTSVYGVTYVGAREQIKRRLEEKGVITD 798
TR|D7LGR7|D7LGR7_ARALL          KDSSKDPESNPTAALAKILITQVDRKLVKQTVMTSVYGVTYVGAREQIKRRLEEKGVITD 797
UPI000859C935                   KDSNKDPESNPTAALAKILINQVDRKLVKQTVMTSVYGVTYVGAREQIKRRLEEKGVISD 789
TR|A0A087GS25|A0A087GS25_ARAAL KDSSKDPESHPTAALAKILINQVDRKLVKQTVMTSVYGVTYVGAREQIKRRLEEKGVITD 787
TR|M4FDQ9|M4FDQ9_BRARP          KDSSKDPESNPTAALAKILINQVDRKLVKQTVMTSVYGVTYVGAREQIKRRLEEKGVITD 786
UPI0006AAEE50                   KDSSKDPESNPTAALAKILINQVDRKLVKQTVMTSVYGVTYVGAREQIKRRLEEKGVITD 786
                                ***.*****:**********.*********************************:*

SP|O24600|RPOT3_ARATH           ERMLFAAACYSAKVTLAALGEIFEAARAIMSWLGDCAKIIASDNHPVRWITPLGLPVVQP 860
TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA ERMLFAAACYSAKVTLAALGEIFEAARAIMSWLGDCAKIIASDNHPVRWITPLGLPVVQP 892
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA ERMLFAAACYSAKVTLAALGEIFEAARAIMSWLGDCAKIIASDNHPVRWITPLGLPVVQP 884
UPI00053999E5                   ERMLFAAACYSAKVTLAALGEIFEAARAIMSWLGDCAKIIASDNHPVRWITPLGLPVVQP 862
UPI00053A373F                   ERMLFAAACYSAKVTLAALGEIFEAARAIMSWLGDCAKIIASDNHPVRWITPLGLPVVQP 862
TR|V4L7M1|V4L7M1_EUTSA          ERMLFSAACYAAKVTLAALGEIFEAARAIMSWLGDCAKIIASDNHPVRWITPLGLPVVQP 860
TR|R0HH96|R0HH96_9BRAS          ERMLFAAACYSAKVTLAALGEIFEAARAIMSWLGDCAKIIASDNHPVRWITPLGLPVVQP 858
TR|D7LGR7|D7LGR7_ARALL          ERMLFAAACYSAKVTLAALGEIFEAARAIMSWLGDCAKIIASDNHPVRWITPLGLPVVQP 857
UPI000859C935                   ERMLFAAACYSAKVTLAALGEIFEAARAIMSWLGDCAKIIASDNHPVRWTTPLGLPVVQP 849
TR|A0A087GS25|A0A087GS25_ARAAL ERMLFAAACYSAKVTLAALGEIFEAARAIMSWLGDCAKIIASDNHPVRWTTPLGLPVVQP 847
TR|M4FDQ9|M4FDQ9_BRARP          ERMLFSAACYSAKVTLAALGEIFEAARAIMSWLGDCAKIIATDNHPVRWTTPLGLPVVQP 846
UPI0006AAEE50                   ERMLFSAACYSAKVTLAALGEIFEAARAIMSWLGDCAKIIATDNHPVRWTTPLGLPVVQP 846
                                *****:****:*******************************:******* **********

SP|O24600|RPOT3_ARATH           YCRSERHLIRTSLQVLALQREGNTVDVRKQRTAFPPNFVHSLDGTHMMMTAVACREAGLN 920
TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA YCRSERHLIRTSLQVLALQREGNTVDVRKQRTAFPPNFVHSLDGTHMMMTAVACREAGLN 952
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA YCRSERHLIRTSLQVLALQREGNTVDVRKQRTAFPPNFVHSLDGTHMMMTAVACREAGLN 944
UPI00053999E5                   YCRSERHLIRTSLQVLALQREGNTVDVRKQRTAFPPNFVHSLDGTHMMMTAVACREAGLN 922
UPI00053A373F                   YCRSERHLIRTSLQVLALQREGNTVDVRKQRTAFPPNFVHSLDGTHMMMTAVACREAGLN 922
TR|V4L7M1|V4L7M1_EUTSA          YCRSERHLIRTSLQVLALQREGNTVDVRKQRTAFPPNFVHSLDGTHMMMTAVACREAGLN 920
TR|R0HH96|R0HH96_9BRAS          YCRSERHLIRTSLQVLALQREGNTVDVRKQRTAFPPNFVHSLDGTHMMMTAVACREAGLN 918
TR|D7LGR7|D7LGR7_ARALL          YCRSERHLIRTSLQVLALQREGNTVDVRKQRTAFPPNFVHSLDGTHMMMTAVACREAGLN 917
UPI000859C935                   YCRSERHLIRTSLQVLALQREGNTVDVRKQRTAFPPNFVHSLDGTHMMMTAVACREAGLN 909
TR|A0A087GS25|A0A087GS25_ARAAL YCRNERHLIRTSLQVLALQREGNTVDVRKQRTAFPPNFVHSLDGTHMMMTAVACREAGLN 907
TR|M4FDQ9|M4FDQ9_BRARP          YCRSERHLIRTSLQVLALQREGNTVDVRKQRTAFPPNFVHSLDGTHMMMTAVACREAGLN 906
UPI0006AAEE50                   YCRSERHLIRTSLQVLALQREGNTVDVRKQRTAFPPNFVHSLDGTHMMMTAVACREAGLN 906
                                ***.************************************************************
```

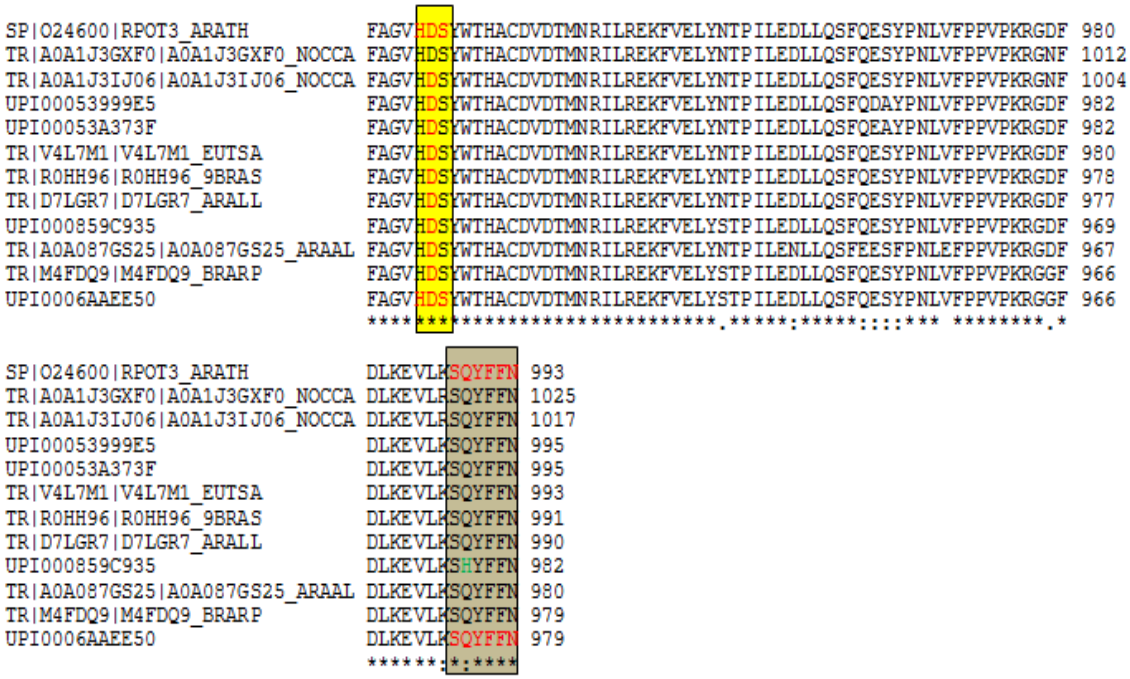


Fig. 3. MSA of various SSU chloroplast RNA polymerases

SP|O24600|RPOT3_ARATH, *Arabidopsis thaliana*

TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA, *Noccaea caerulescens*

TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA, *Noccaea caerulescens* (Alpine penny-cress) (*Thlaspi caerulescens*)

UPI00053999E5, *Capsella rubella*

UPI00053A373F, *Arabidopsis lyrata*

TR|V4L7M1|V4L7M1_EUTSA, *Eutrema salsugineum* (Saltwater cress)

TR|R0HH96|R0HH96_9BRAS,*Capsella rubella*

TR|D7LGR7|D7LGR7_ARALL,*Arabidopsis lyrata* subsp. lyrata (Lyre-leaved rock-cress)

UPI000859C935, *Raphanus sativus* (Radish)

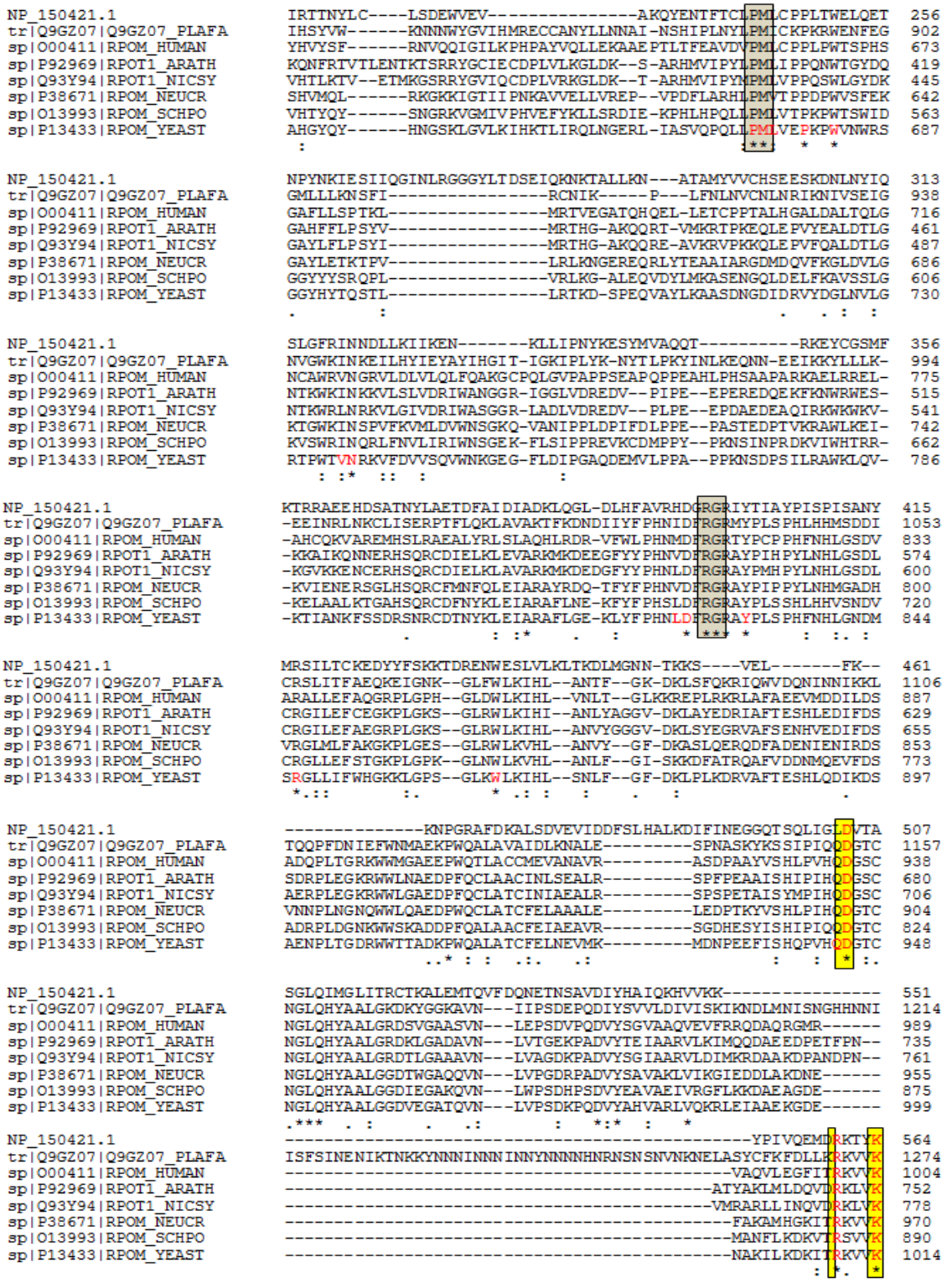
TR|A0A087GS25|A0A087GS25_ARAAL, *Arabis alpina* (Alpine rock cress)

TR|M4FDQ9|M4FDQ9_BRARP, *Brassica rapa subsp. pekinensis* (Chinese cabbage)

UPI0006AAEE50, *Brassica napus* (Rape)

Fig. 4 shows the MSA and conserved motifs in SSU mitochondrial RNA polymerases from different sources like plants and fungi. It was found that the N terminal and C- terminal regions are devoid of many conserved motifs. However, the middle region towards the C terminal region shows strong alignment and showing many conserved motifs among them. The catalytic, template and substrate binding motifs are highlighted. The YG gate keeper motif and the catalytic K is strictly conserved (including distance conservation) in SSU mitochondrial RNA polymerases also This strongly suggests that all these RNA polymerases might be using same set of amino acids for template, substrate binding and catalysis. The immediate downstream amino acid in DNA polymerases is usually a G or A [4] but in these RNA polymerases, it is mostly a Q. like chloroplast RNA polymerases (The brown alga uses G and the *Schizosaccharomyces* uses P). Interestingly, here also the 4th amino acid, R, is the invariant amino acid suggesting a possible role in substrate binding and /or catalysis.

CLUSTAL O(1.2.4) MSA and conserved motifs in mitochondrial RNA polymerases from plant and fungi.



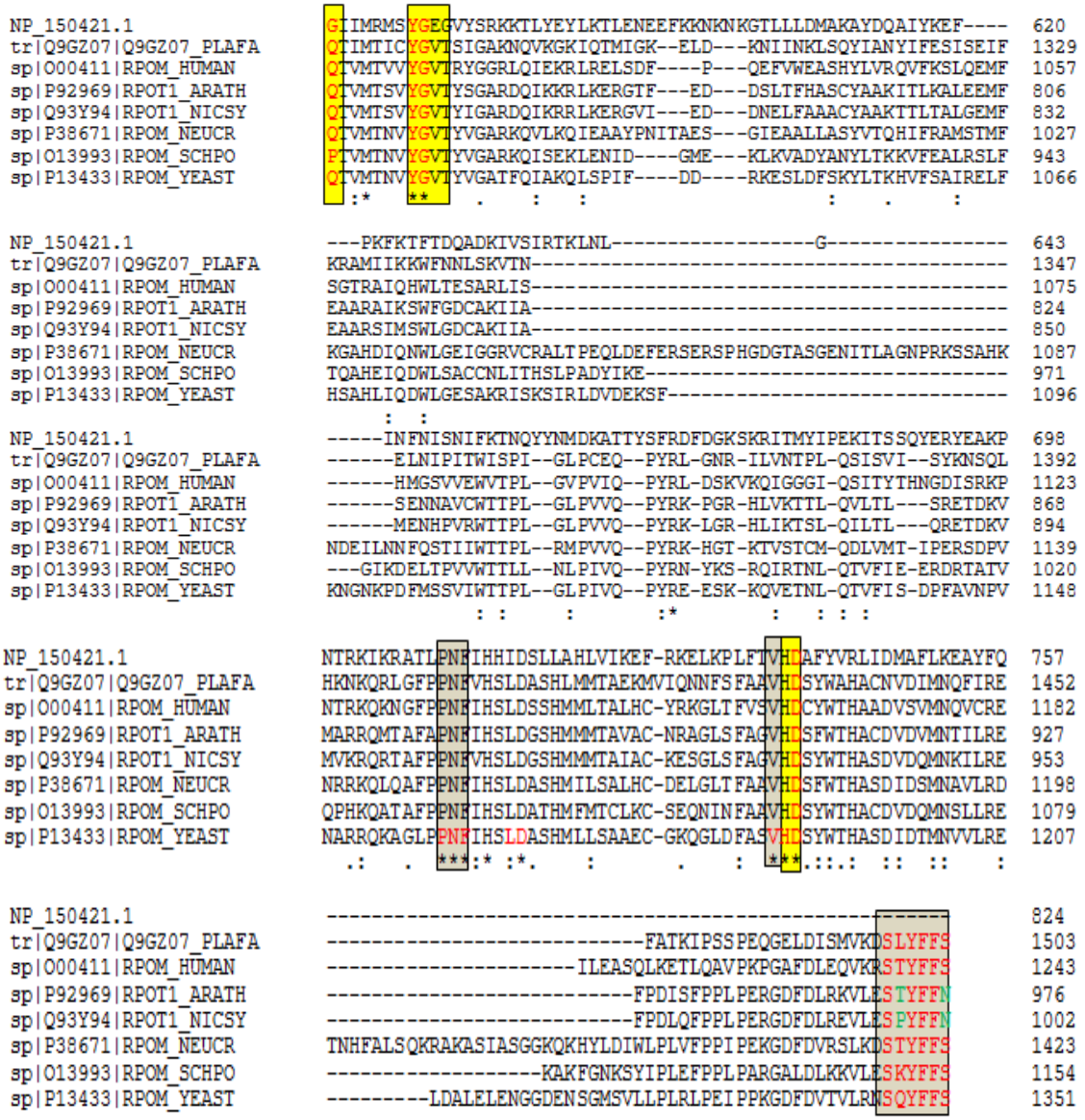


Fig. 4. MSA of mitochondrial RNA polymerases from plants and fungi

*The N terminal motifs are different from plant RNA polymerases and the amino acids are marked in green

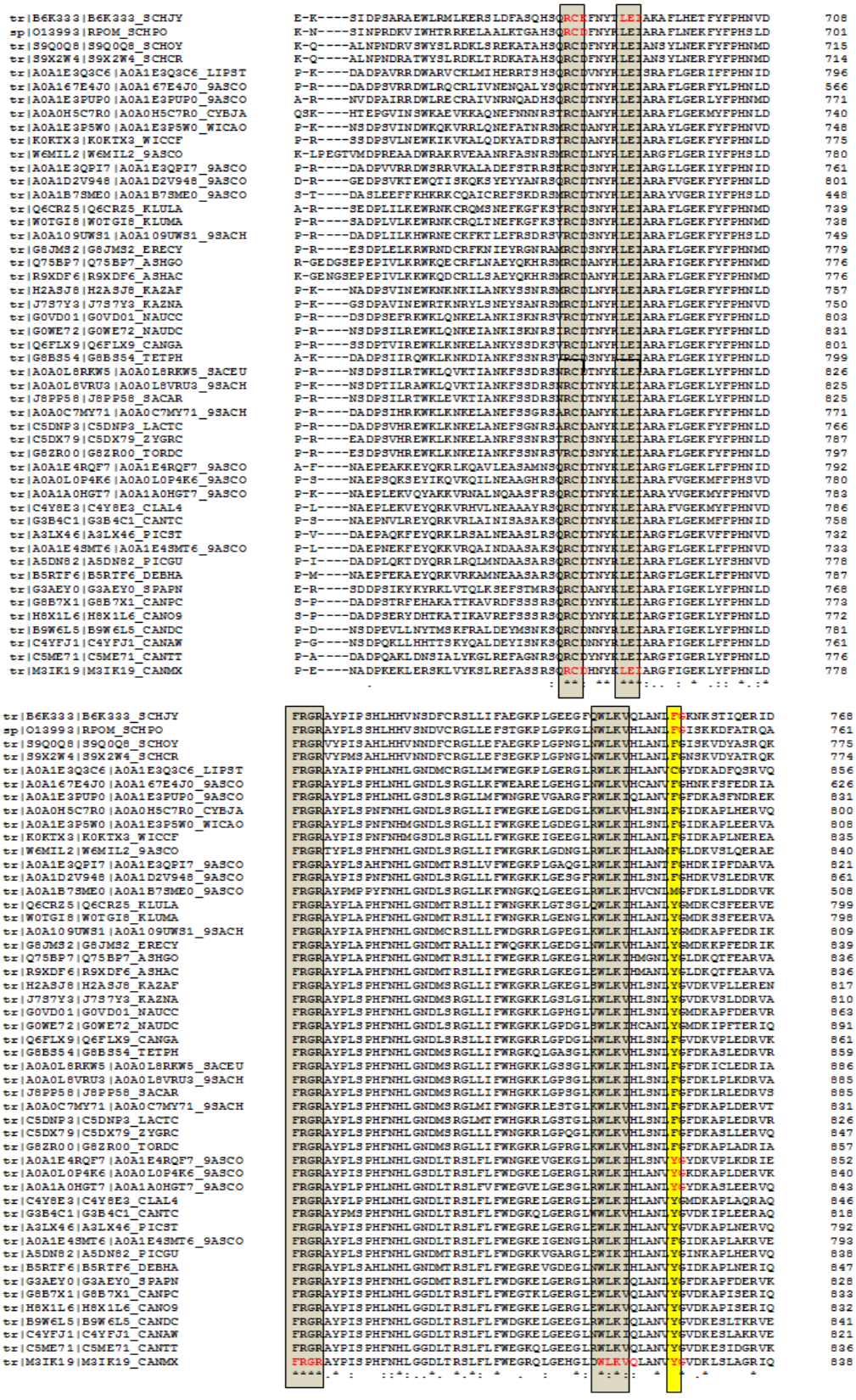
NP_150421.1, *Pylaiella littoralis* (Sea weed, brown alga)
tr|Q9GZ07|Q9GZ07_PLAFA, *Plasmodium falciparum* (Malarial parasite)
sp|O00411|RPOM_HUMAN, *Homo sapiens*
sp|P92969|RPOT1_ARATH, *Arabidopsis thaliana*
sp|Q93Y94|RPOT1_NICSY, *Nicotiana sylvestris*
sp|P38671|RPOM_NEUCR, *Neurospora crassa*
sp|O13993|RPOM_SCHPO, *Schizosaccharomyces pombe*
sp|P13433|RPOM_YEAST, *Saccharomyces cerevisiae*

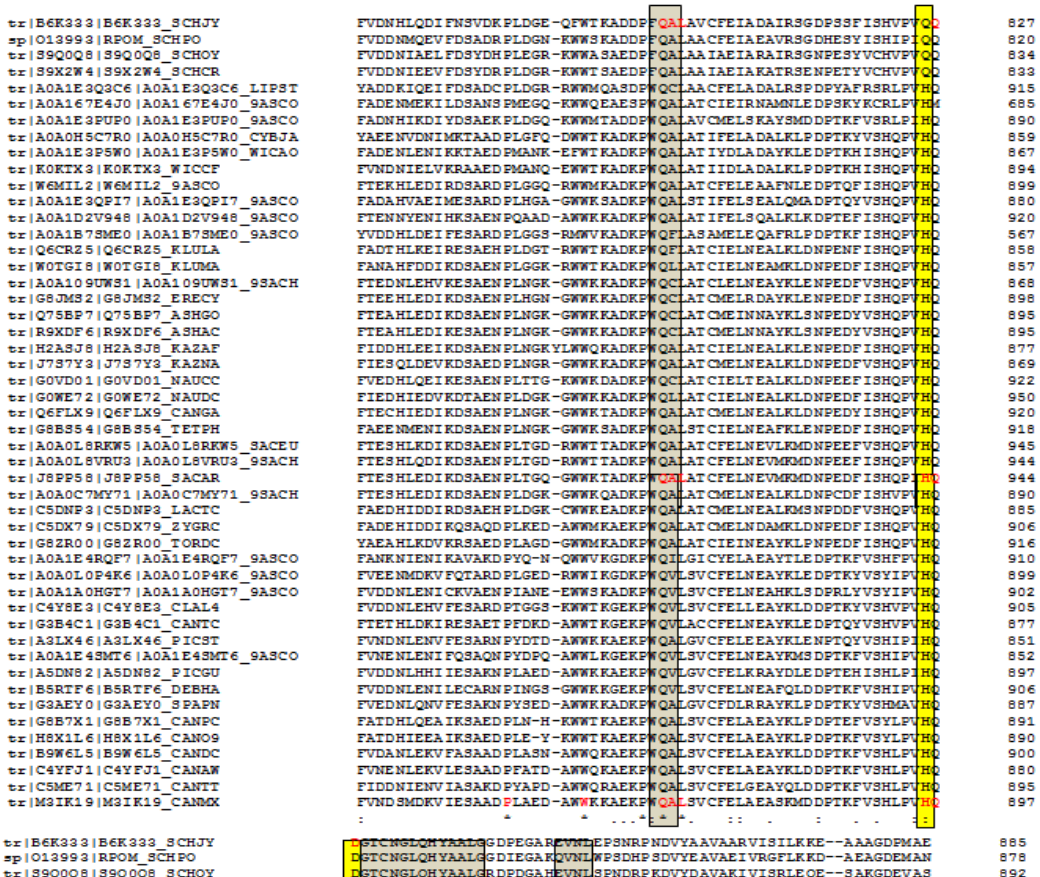
Fig. 5 shows the MSA and conserved motifs in SSU RNA polymerases exclusively from fungal mitochondria. It is clear that there are no highly conserved motifs in the N-terminal and C-terminal regions. However, the middle region towards the C terminal shows large regions of

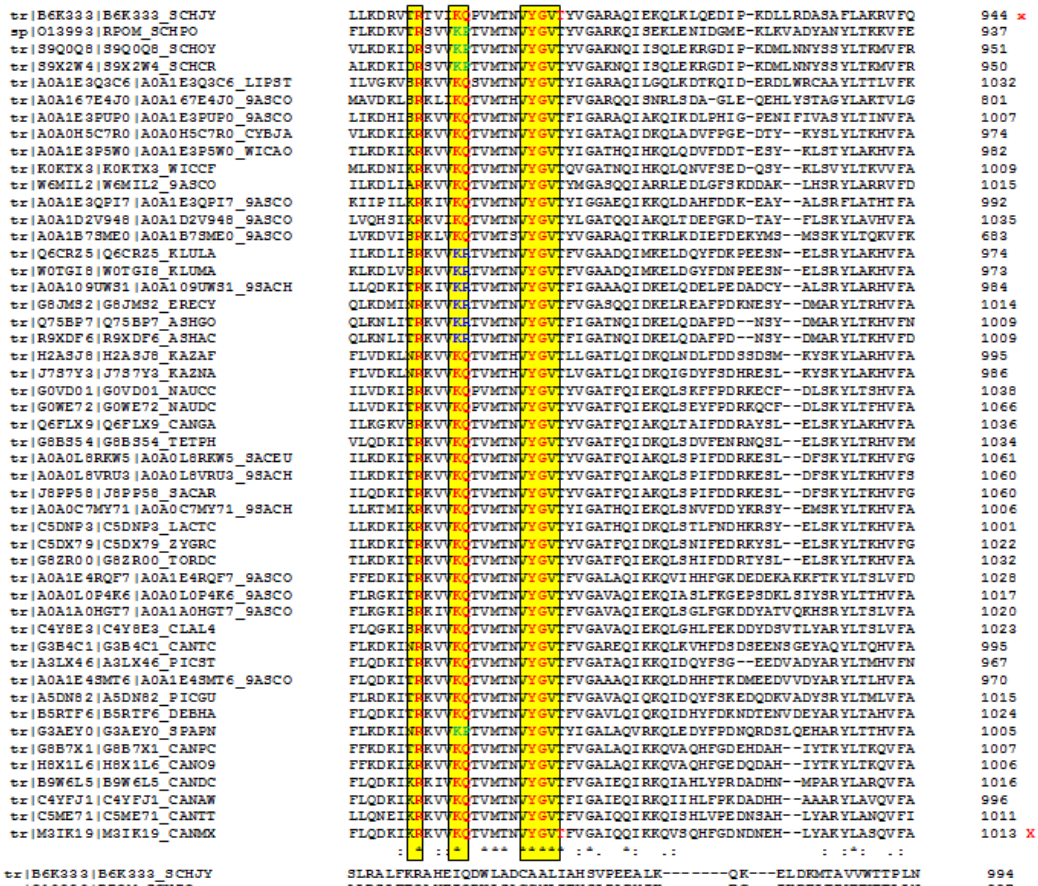
conservation including the catalytic K, YG pair and an invariant R. The catalytic, template and substrate binding motifs are highlighted. The YG gate keeper motif and the catalytic K is strictly conserved (including distance conservation) in all DNA dependent RNA polymerases from fungal mitochondria. This strongly suggests that all these RNA polymerases might be using same set of amino acids for template, substrate binding and catalysis. The immediate downstream amino acid in DNA polymerases is usually a G or A [4] but in these out of 49 sequences analyzed, 39 RNA polymerases uses Q. (Six use R and three use P and one uses G). Interestingly, here also the 4th amino acid, R, is the invariant amino acid in all the 49 sequences, suggesting an important role in substrate binding and /or catalysis. All these fungal mitochondrial RNA polymerases end in '**SxYFFS**,' and its role is not known as of now. Identical sequences were seen in *Plasmodium* and human mitochondrial polymerases but the plant mitochondrial polymerases slightly vary (Fig. 5). This 6-amino acid sequence is found in RNA binding protein of the fungus, *Ustilago maydis* and primarily involve in RNA transports [6].

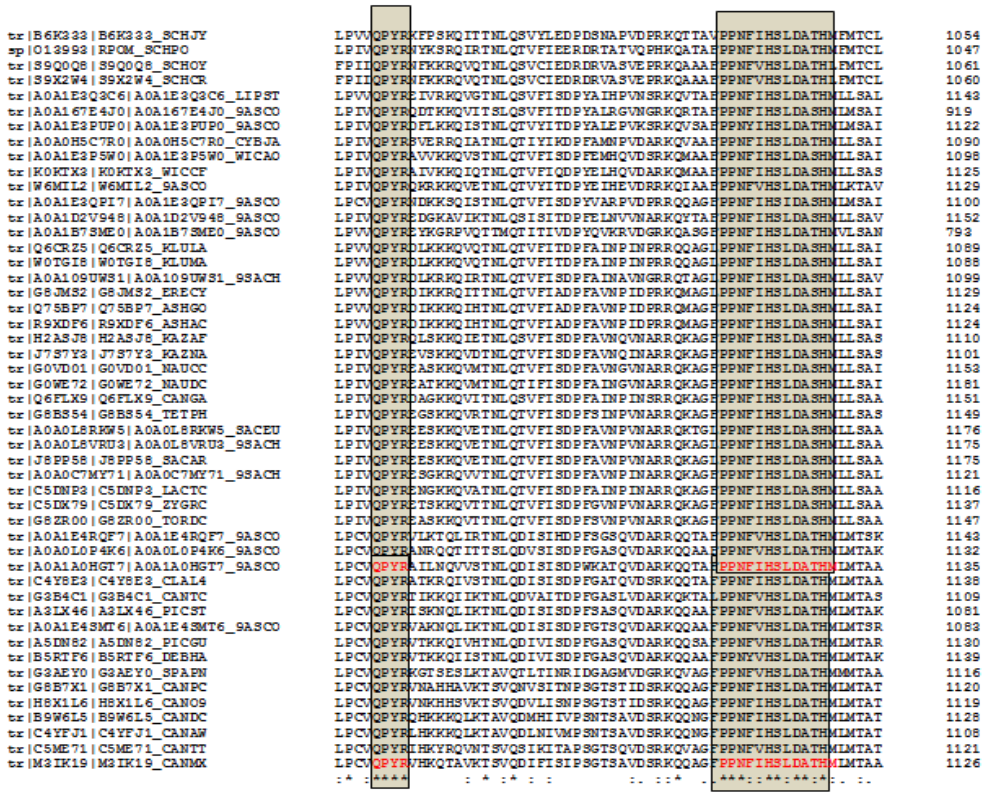
CLUSTAL O(1.2.4) MSA various SSU mitochondrial RNA polymerases from fungi

```
tr|B6K333|B6K333_SCHJY          AASDSGQLKDLLEGLTALGNVGWKVNRKVYDMLVKIWNTGESFLSIPSAN-TTLDIQEMP   653
sp|O13993|RPOM_SCHPO            KASENGQLDELFKAVSSLGKVSWRINQRLFNVLIRIWNSGEKFLSIPPRE-VKCDMPPYP   646
tr|S9Q0Q8|S9Q0Q8_SCHOY          EASHKGHLKKIYNALSALGDVDWRINRFTFDVIVKIWNSGEGMLSIPPKN-VEVNLPPYP   660
tr|S9X2W4|S9X2W4_SCHCR          EASRKGHLKKVYGALSALGDVSWRINRFTFDVIVKIWNSGEGMLSIPPKN-VEVNLPPYP   659
tr|A0A1E3Q3C6|A0A1E3Q3C6_LIPST  EACRRNDLESVYEGLDVLGSAAWIINTRVFEVLAKVWNTGEEFLEIPTRYEGDINFPLEP   741
tr|A0A167E4J0|A0A167E4J0_9ASCO  AASDRGTLDQVYEGLNVLGNTAWTINKDIFEVILKVWNTGEEFLEIPRRVEIQPELPPAP   511
tr|A0A1E3PUP0|A0A1E3PUP0_9ASCO  EASKRGAMNEVFEGLNVLGNTSWTINKDIFKVVMQLWNTGKPFLDIPGIS-DDVELPTEP   716
tr|A0A0H5C7R0|A0A0H5C7R0_CYBJA  ASSDK--IGKVYEGLNVLGSTAWTVNERVFNVMSQVWNTGEEFLDIPKLQ-DKIELPPQP   684
tr|A0A1E3P5W0|A0A1E3P5W0_WICAO  ASTDR--LDLVYEGLNVLGETAWTVNERVYKVMSEIWNTGETFLDIPKDQ-EEMNVPDAP   693
tr|K0KTX3|K0KTX3_WICCF          ASSDK--LDKVYDGLNVLGNTAWTVNERVFKVMSKIWNTGDEFLDIPKAH-ESLELPEVP   720
tr|W6MIL2|W6MIL2_9ASCO          AASKRGDLDKVFKGLNVLGNTQWTPNKRILEIVTQVWNSGEEMLEIPAHI-SELKLPDPP   721
tr|A0A1E3QPI7|A0A1E3QPI7_9ASCO  AASENGSLASVYKGLTVLGDTPWTVNRKIYDIVSQVWNTGESFLDIAGVQ-DELELPPPP   706
tr|A0A1D2V948|A0A1D2V948_9ASCO  AASENGDLEGVYDGLNVLGVTAWTINKKILDLVTEAWNSGEEFLDIPKHE-KELILPPKP   746
tr|A0A1B7SME0|A0A1B7SME0_9ASCO  TAAIKGKMDTVLQALNNLGSTAWTVNKEVLKVMIQVWNTGEEFLDIPKFR-ETLNLPPQP   393
tr|Q6CRZ5|Q6CRZ5_KLULA          AVSEQGSIDNVYEGLNVLGNTPWTINEKLFNIVSEVWNTGEGFLDIPPVL-DNFEFPPQP   684
tr|W0TGI8|W0TGI8_KLUMA          AVSEQGSIQNVYEGLNVLGNTAWTINEKLFNVISEVWNTGEEFLDIPPLQ-DKLELPAQP   683
tr|A0A109UW31|A0A109UW31_9SACH  AVSDK--LDNVYKGMNVLGDTPWTVNKTMLNIISTIWNSGEGFLDIPQK-EDVELPPRP   694
tr|G8JMS2|G8JMS2_ERECY          AVTGKGAVNNIYQGLNVLGDTAWTVNKPLFTILSKIWNSGEEYLDIPPQE-DDCQLPPKP   724
tr|Q75BP7|Q75BP7_ASHGO          AVTGKGVVQNVYKGLNVLGETAWTVNKNMLHIISKVWNSGTEFLDIPPQQ-DELLLPAKP   717
tr|R9XDF6|R9XDF6_ASHAC          AVTGKGAVQNVYNGLNVLGETAWTVNKNMLHIVSKVWNSGSEFLDIPPQQ-DELLLPPRP   717
tr|H2ASJ8|H2ASJ8_KAZAF          AASNVQALDKVYDGLNVLGDTAWTVNRRIFEIISKVWNKGGSFLEIPGIQ-QESELPLPP   702
tr|J7S7Y3|J7S7Y3_KAZNA          AASNSNAIDKVYDGLNVLGNTAWTVNRNIFNVVSKVWNEGKPFLEIPGTQ-EILELLPPP   695
tr|G0VD01|G0VD01_NAUCC          AVSDADVIPDVYTGLNALGETAWTINKKVFDVISQVWNTGDEFLEIPEIQ-HDMNLLPPP   748
tr|G0WE72|G0WE72_NAUDC          AVSDKEAIDDVYNGLNALGKTAWTVNKRLFEIISKVWNEGIEFLDIPQIQ-YDMQLIPPP   776
tr|Q6FLX9|Q6FLX9_CANGA          AVSDSGAIDKIYDGLNVLGETAWTVNRKVFDVVSKVWNEGTNFLEIPGIQ-DELKLLPPP   746
tr|G8B854|G8B854_TETPH          AVSDAGAIDRVYHGLNVLGSTPWTVNRKIFNVISEVWNTGKPFLDIPGIQ-NTLNLLPPP   744
tr|A0A0L8RKW5|A0A0L8RKW5_SACEU  AASENGDIDRVYDGLNVLGDTPWTVNRKVFDVVSQVWNNGEGFLDIPSSQ-DEMILPPAP   771
tr|A0A0L8VRU3|A0A0L8VRU3_9SACH  AASDNGDIDRVYDGLNVLGRTPWTVNRKVFDVVSQVWNKGEGFLDIPGAQ-DEMVLPPAP   770
tr|J8PP58|J8PP58_SACAR          AASENGDIDRVYDGLNVLGNTPWTVNRKVFDVVSQVWNKREGFLDIPGAQ-DEMVLPPAP   770
tr|A0A0C7MY71|A0A0C7MY71_9SACH  AVSDSGAIDNVYHGLNVLGETPWTVNRKMFDIMSHVWNTGKAFLDIPGNQ-SKLILTPQP   716
tr|C5DNP3|C5DNP3_LACTC          AVSDAGAIDNVYRGLNVLGETPWTVNRKMFNVMSQVWNSGKPFLDIPGIQ-EKLTLTPQP   711
tr|C5DX79|C5DX79_ZYGRC          AVSDATAINTVYDGLNVLGETPWTVNRPLFEVMSKVWNSGKPFLDIPGVQ-EKLTLLPPP   732
tr|G8ZR00|G8ZR00_TORDC          AVSNAGAIDTVYQGLNVLGETAWTVNRRVFDVMSEVWNSGKDFLEIPGIQ-EKLDLLPAP   742
tr|A0A1E4RQF7|A0A1E4RQF7_9ASCO  ASSN--RISGVYDGLNVLGETPWTINRKVFEVITHYWNTGKQFLDIPPVS-EEINYPEPV   737
tr|A0A0L0P4K6|A0A0L0P4K6_9ASCO  AASDAHNLDEVYKGLNILGETAWTVNSKVLDVISRCWNTGEQFLDIPPVA-EEPKLPPPL   725
tr|A0A1A0HGT7|A0A1A0HGT7_9ASCO  AASDSGRLDGVYAGLNVLGKTAWSVNAKVLEVISHYWNTGNEFLDIPPIV-EEPKLPEPL   728
tr|C4Y8E3|C4Y8E3_CLAL4          AASEANNLDDVYRGLNVLGHTPWTINAKVLEVISQYWNTGEAFLDIPPVV-DEPELPPPL   731
tr|G3B4C1|G3B4C1_CANTC          AASQRGDLDKVYDGLNVLGNTPWTVNRKVLEVMTKFWNTGDEFLDIPPVV-DEPELPEPV   703
tr|A3LX46|A3LX46_PICST          KAADLGNLNEVYDGLNVLGKTPWTVNRRVFEIITRYWNSGEEFLDIPPIM-EEPNLPEPL   677
tr|A0A1E4SMT6|A0A1E4SMT6_9ASCO  AASEMGNLDEIYQGLNVLGNTAWTVNRRVLDVITKHWNTGKEFLDIPPII-EEPNLPKAI   678
tr|A5DN82|A5DN82_PICGU          AASDMGNLDQVYEGLNVLGETCWTINHEVFDVISHYWNSGEKFLDIPSIV-EELDIPPPP   723
tr|B5RTF6|B5RTF6_DEBHA          AASDLDNLE-IYDGLNVLGDTAWTVNRNVFNIISKFWNTGEKFLDIPPIV-EEPELPKQL   732
tr|G3AEY0|G3AEY0_SPAPN          TAARNGNLDQVYAGLNVLGNTAWTVNKDIFKVISHYWNTGKEFLDIPGVL-DKPILPPKP   713
tr|G8B7X1|G8B7X1_CANPC          AAAKRGNLKEVFDGLNVLGTTAWTINKKVFDIISHYWNKGEEFLSIPPVL-ETAQFPAKL   718
tr|H8X1L6|H8X1L6_CANO9          AAAKRGNLQEVFDGLNVLGTTAWTINKKVFDVISHYWNKGDEFLTIPPVL-ESAQFPPKL   717
tr|B9W6L5|B9W6L5_CANDC          AANERGNLKAVFEGLNVLGKTAWTVNKKVFNVISHYWNSGEEFLSIPPVS-EEPRLPKDV   726
tr|C4YFJ1|C4YFJ1_CANAW          AANERGNLTGVFEGLNVLGNTAWTVNKKVFNVISHYWNSGEEFLSIPPVS-EEPQLPQDI   706
tr|C5ME71|C5ME71_CANTT          AASKRGNLDQVFDGLNVLGNTAWTVNKRVFDVISHYWNKGEEFLSIPPVM-EEARLPEKL   721
tr|M3IK19|M3IK19_CANMX          AASKRGNLDEVFKGLNVLGSTPWTVNSKVFDVISHYWNIGEEFLSIPPVL-NEAKFPSEL   723
                                                                                              
```









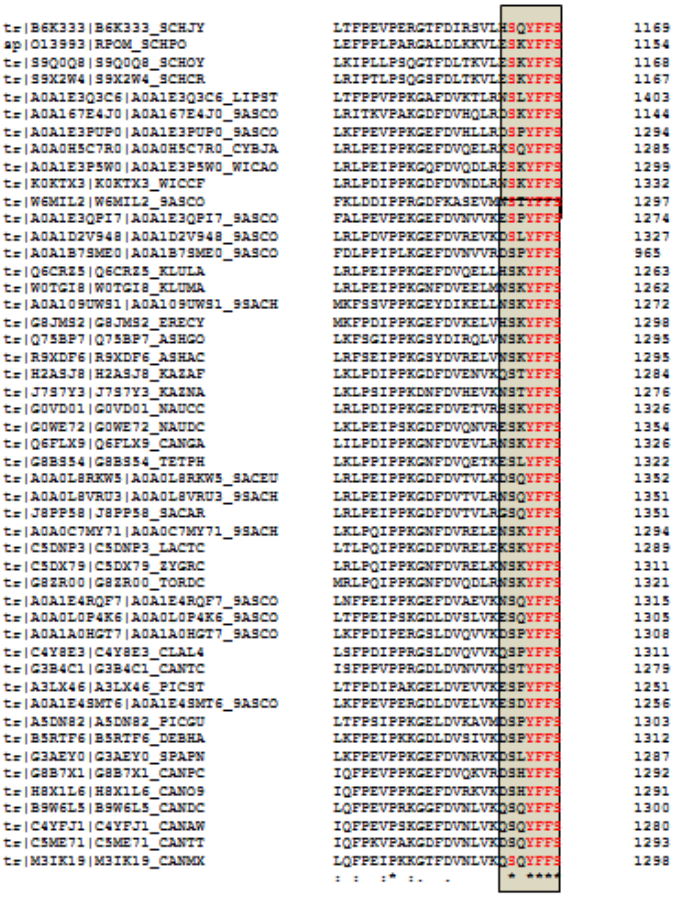


Fig. 5. MSA of various SSU RNA polymerases from various fungal mitochondria

*All the mitochondrial RNA polymerases end in SxYFFS

tr|B6K333|B6K333_SCHJY, *Schizosaccharomyces japonicus*
sp|O13993|RPOM_SCHPO, *Schizosaccharomyces pombe*
tr|S9Q0Q8|S9Q0Q8_SCHOY, *Schizosaccharomyces octosporus*
tr|S9X2W4|S9X2W4_SCHCR, *Schizosaccharomyces cryophilus*
tr|A0A1E3Q3C6|A0A1E3Q3C6_LIPST, *Lipomyces starkeyi*
tr|A0A167E4J0|A0A167E4J0_9ASCO, *Sugiyamaella lignohabitans*
tr|A0A1E3PUP0|A0A1E3PUP0_9ASCO, *Nadsonia fulvescens var. elongata*
tr|A0A0H5C7R0|A0A0H5C7R0_CYBJA, *Cyberlindnera jadinii*
tr|A0A1E3P5W0|A0A1E3P5W0_WICAO, *Wickerhamomyces anomalus*
tr|K0KTX3|K0KTX3_WICCF, *Wickerhamomyces ciferrii*
tr|W6MIL2|W6MIL2_9ASCO, *Kuraishia capsulata*
tr|A0A1E3QPI7|A0A1E3QPI7_9ASCO, *Babjeviella inositovora*
tr|A0A1D2V948|A0A1D2V948_9ASCO, *Ascoidea rubescens*
tr|A0A1B7SME0|A0A1B7SME0_9ASCO, *Ogataea polymorpha*
tr|Q6CRZ5|Q6CRZ5_KLULA, *Kluyveromyces lactis*
tr|W0TGI8|W0TGI8_KLUMA, *Kluyveromyces marxianus*
tr|A0A109UWS1|A0A109UWS1_9SACH, *Eremothecium sinecaudum*
tr|G8JMS2|G8JMS2_ERECY, *Eremothecium cymbalariae*
tr|Q75BP7|Q75BP7_ASHGO, *Ashbya gossypii*
tr|R9XDF6|R9XDF6_ASHAC, *Ashbya aceri*
tr|H2ASJ8|H2ASJ8_KAZAF, *Kazachstania africana*
tr|J7S7Y3|J7S7Y3_KAZNA, *Schizosaccharomyces octosporus*
tr|G0VD01|G0VD01_NAUCC, *Naumovozyma castellii*

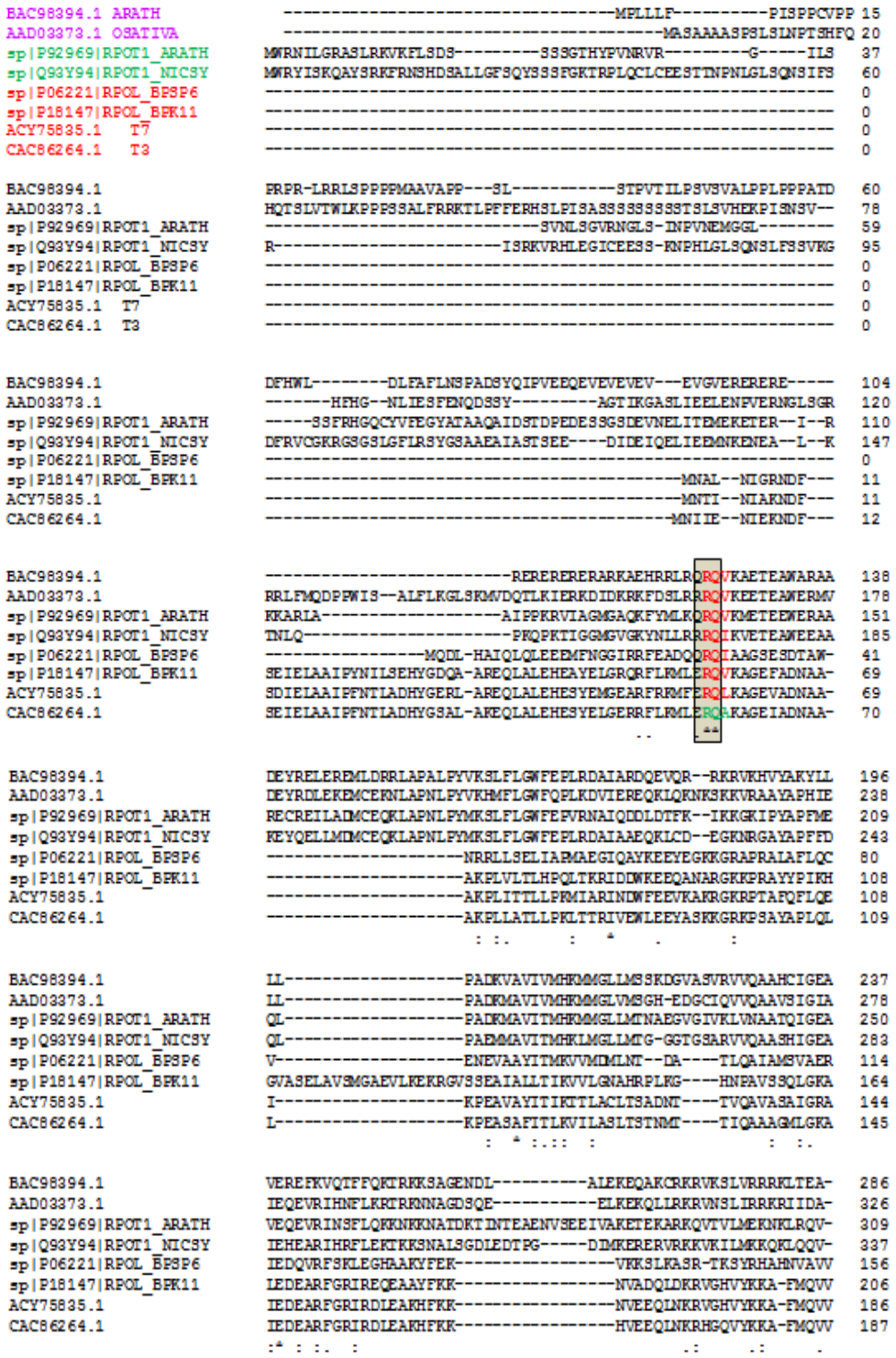
tr|G0WE72|G0WE72_NAUDC,*Naumovozyma dairenensis*
tr|Q6FLX9|Q6FLX9_CANGA, *Candida glabrata*
tr|G8BS54|G8BS54_TETPH, *Tetrapisispora phaffii*
tr|A0A0L8RKW5|A0A0L8RKW5_SACEU, *Saccharomyces eubayanus*
tr|A0A0L8VRU3|A0A0L8VRU3_9SACH, *Saccharomyces sp. 'boulardii'*
tr|J8PP58|J8PP58_SACAR, *Saccharomyces arboricola*
tr|A0A0C7MY71|A0A0C7MY71_9SACH, *Lachancea lanzarotensis*
tr|C5DNP3|C5DNP3_LACTC, *Lachancea thermotolerans*
tr|C5DX79|C5DX79_ZYGRC, Zygosaccharomyces rouxii
tr|G8ZR00|G8ZR00_TORDC, *Torulaspora delbrueckii*
tr|A0A1E4RQF7|A0A1E4RQF7_9ASCO, *Hyphopichia burtonii*
tr|A0A0L0P4K6|A0A0L0P4K6_9ASCO, *Candida auris*
tr|A0A1A0HGT7|A0A1A0HGT7_9ASCO, *Metschnikowia bicuspidata var. bicuspidata*
tr|C4Y8E3|C4Y8E3_CLAL4, *Clavispora lusitaniae*
tr|G3B4C1|G3B4C1_CANTC, *Candida tenuis*
tr|A3LX46|A3LX46_PICST, *Scheffersomyces stipitis*
tr|A0A1E4SMT6|A0A1E4SMT6_9ASCO, *Candida tanzawaensis*
tr|A5DN82|A5DN82_PICGU, *Meyerozyma guilliermondii*
tr|B5RTF6|B5RTF6_DEBHA, *Debaryomyces hansenii*
tr|G3AEY0|G3AEY0_SPAPN, *Spathaspora passalidarum*
tr|G8B7X1|G8B7X1_CANPC, *Candida parapsilosis*
tr|H8X1L6|H8X1L6_CANO9, *Candida orthopsilosis*
tr|B9W6L5|B9W6L5_CANDC, Candida dubliniensis
tr|C4YFJ1|C4YFJ1_CANAW, *Candida albicans*
tr|C5ME71|C5ME71_CANTT, *Candida tropicalis*
tr|M3IK19|M3IK19_CANMX, *Candida maltosa*

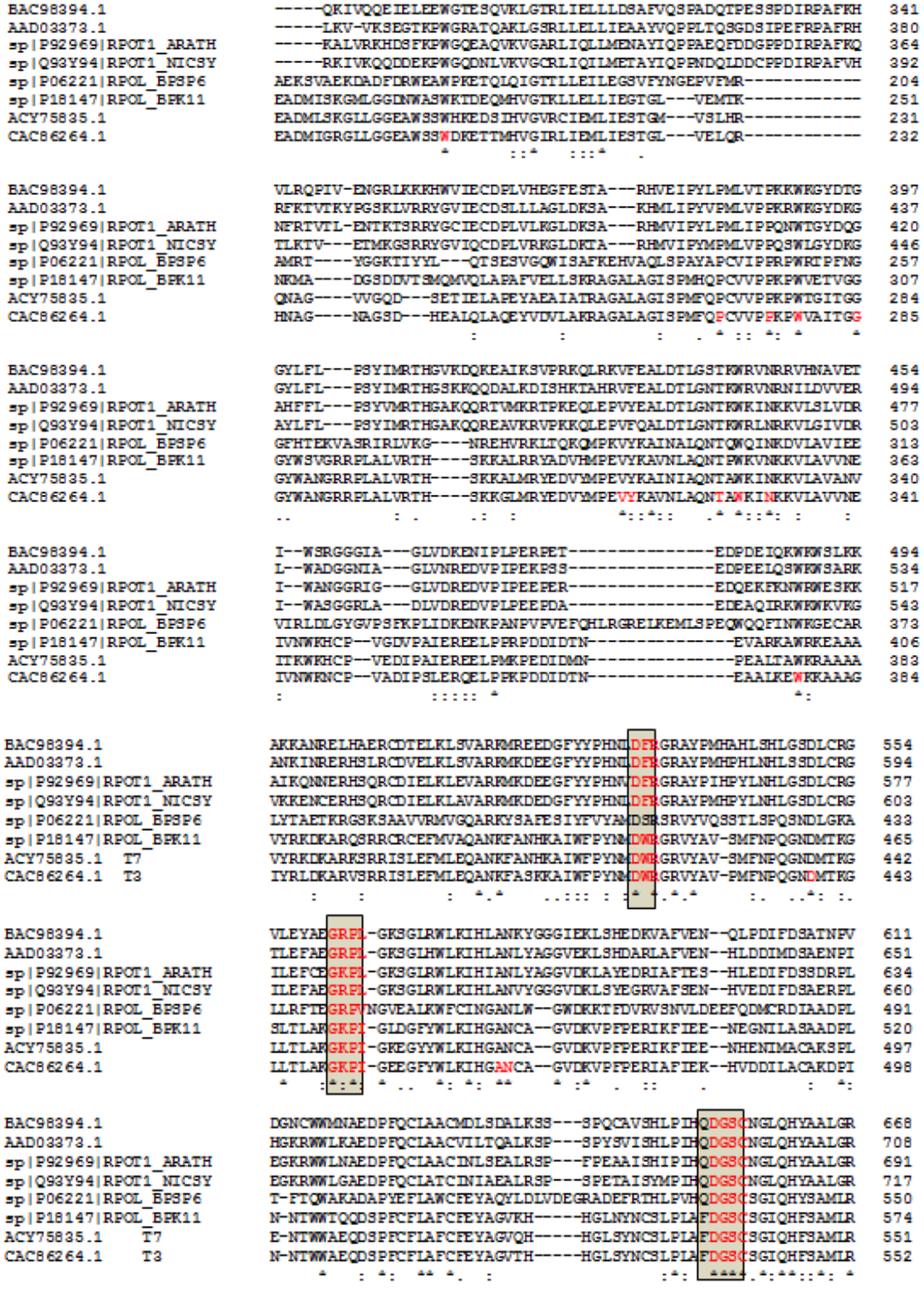
In the viral, chloroplast and mitochondrial RNA polymerases the last 6 amino acids at N-terminal ends are highly conserved, suggesting a possible role for this motif in transcription cycle. Whether it possibly involves in a Rho-dependent or independent termination process to offload the nascent RNA at the termination site is to be elucidated. The consensus motif in viral polymerases is **-SDFAFA.** Peptide search analysis shows this motif is also found in RNA binding protein of the fungus, *Ustilago maydis* and primarily involves in RNA transports [6] and in poyA binding protein where it is implicated in both mRNA cleavage and polyadenylation in the nucleus. In chloroplast and mitochondrial polymerases the consensus sequence is -**SxYFFS. (**The -YPPS tetrad on peptide search analysis is found in ATP-dependent DNA helicase and an ATP-dependent, dual-direction single-stranded exonuclease. This tetrad is also found in plant and human transcriptional activators). Therefore, it is probably involved in termination, i.e., transcript cleavage process.

The metal binding sites are also highly conserved among all these SSU RNA polymerases which are highlighted in yellow. Usually, a D in QD and a D in HDS are found to be involved in binding to Mg^{2+} and in 'NTP charge shielding' and found in all these RNA polymerases.

Fig. 6 shows the MSA of a 'mix and match' analysis and shows the conserved motifs in all the three different categories of SSU RNA polymerases, viz. viruses (4), chloroplasts (2) and mitochondria (2). Such an analysis may narrow down the only motifs common among them that may be essential for substrate binding and catalysis. More conserved regions are seen towards the C-terminal regions. The catalytic, template and substrate binding motifs are highlighted. The C terminal region shows conservation in the catalytic K, YG pair and an invariant R among them as expected. The YG gate keeper motif and the catalytic K is strictly conserved (including distance conservation) in DNA dependent RNA polymerases from all the three different sources. This is in accordance with the DNA polymerases data, reported by Palanivelu [4]. This strongly suggests that the DNA and RNA polymerases might be using the same set of amino acids for template, substrate binding and catalysis.

CLUSTAL O(1.2.4) MSA - 4 viral, 2 Mitochondria RNA polymerases




BAC98394.1
AAD03373.1
sp|P92969|RPOT1_ARATH
sp|Q93Y94|RPOT1_NICSY
sp|P06221|RPOL_BPSP6
sp|P18147|RPOL_BPK11
ACY75835.1
CAC86264.1
T7
T3

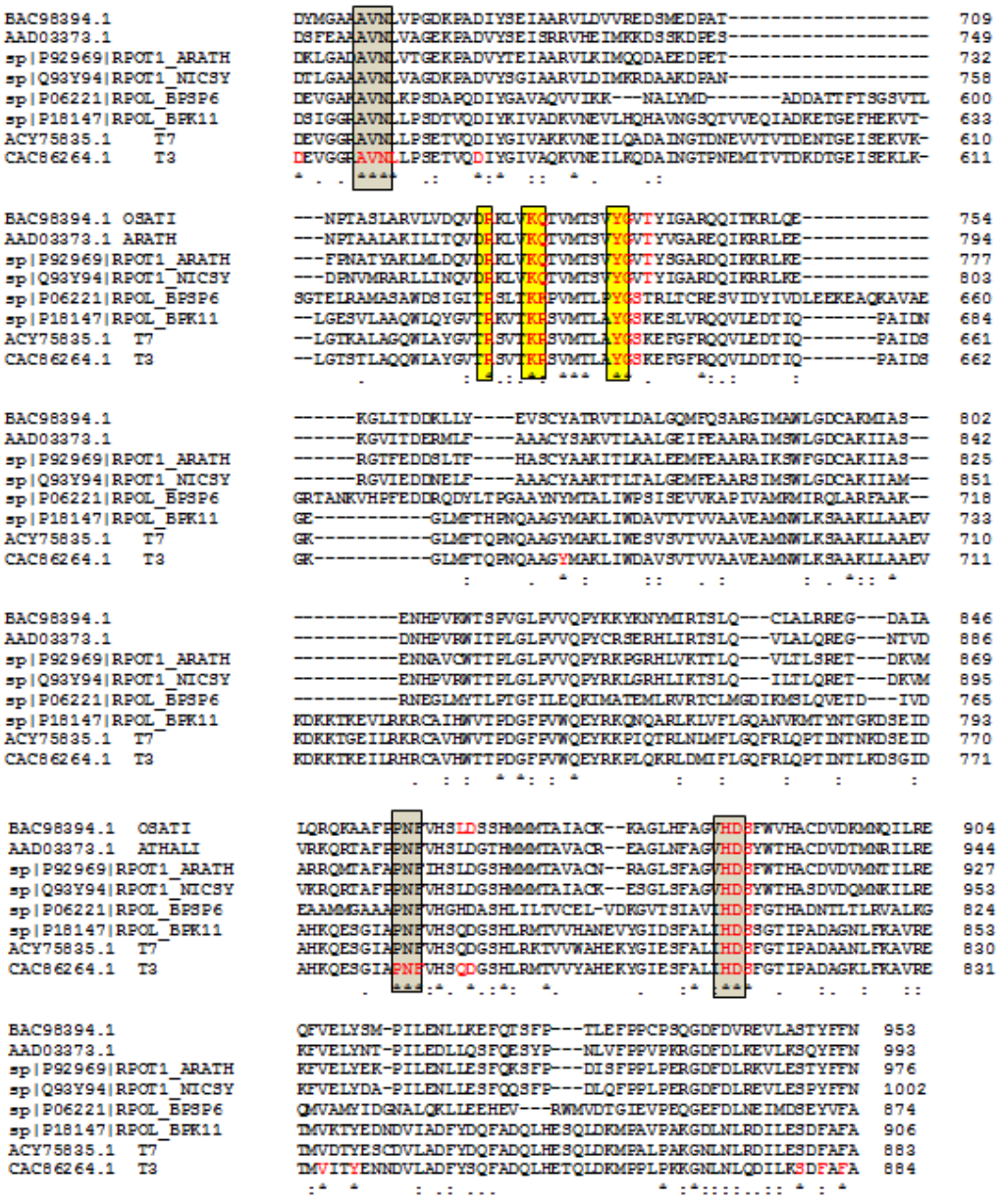


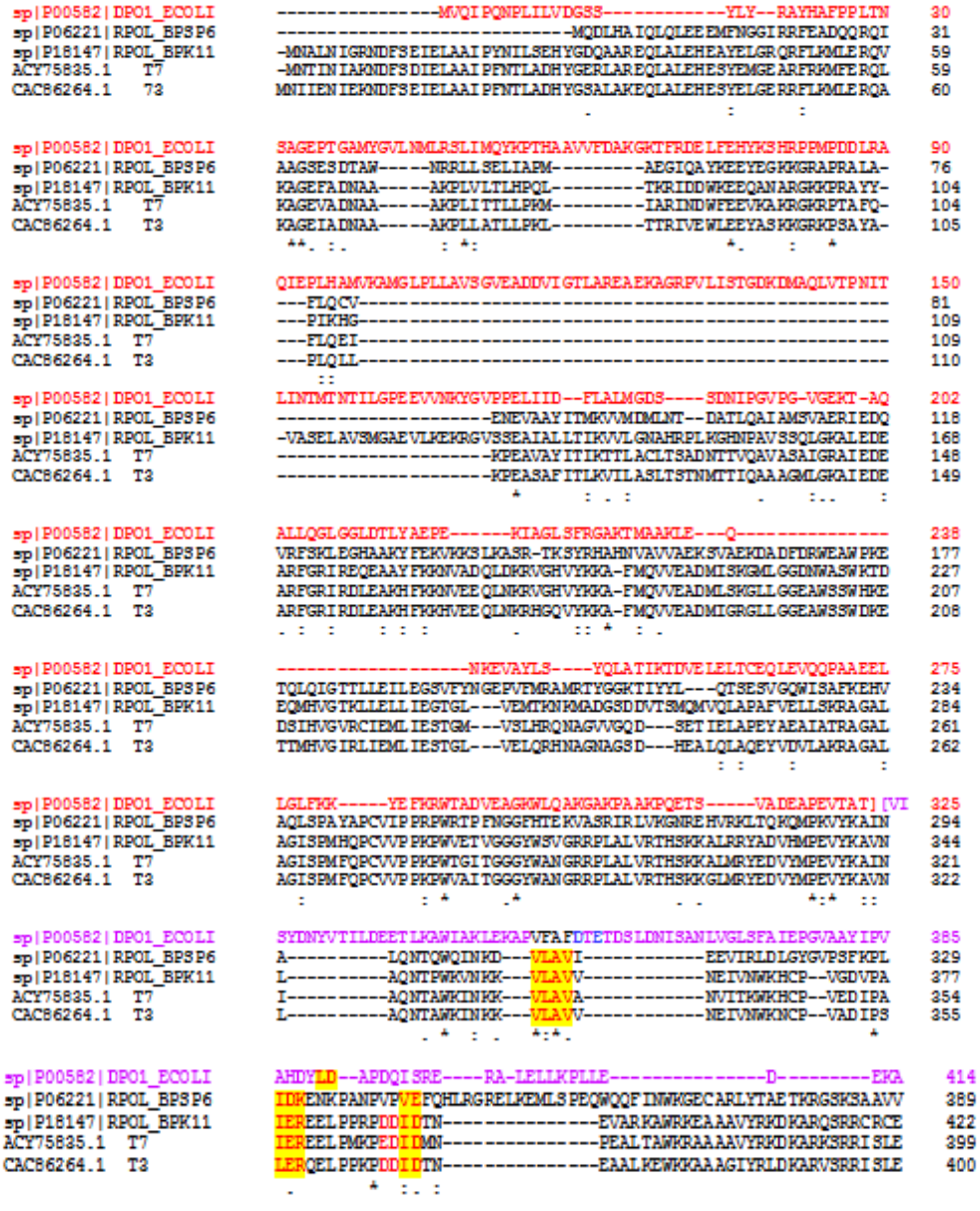
Fig. 6. MSA of T3, T7, K11, SP6 with SSU mitochondrial and chloroplast RNA polymerases

BAC98394.1, *Oryza sativa*, Japonica group (Chloroplast)
AAD03373.1, *Arabidopsis thaliana* (Chloroplast)
sp|P92969|RPOT1_ARATH, *Arabidopsis thaliana* (Mitochondria)
sp|Q93Y94|RPOT1_NICSY, *Nicotiana sylvestris* (Mitochondria)
sp|P06221|RPOL_BPSP6, Enterobacteria phage SP6
sp|P18147|RPOL_BPK11, Enterobacteria phage K11
ACY75835.1 T7, Enterobacteria phage T7
CAC86264.1 T3, Enterobacteria phage T3

Fig. 7 shows the results of MSA of a 'mix and match' analysis between DNA and RNA polymerases, viz. the *E. coli* DNA polymerase I and the RNA polymerases. In this analysis, the *E. coli* DNA polymerase I and the 4 viral RNA polymerases are mixed and aligned to identify

the conserved motifs common to both of them. Such an analysis may narrow down the motifs common among DNA and RNA polymerases that may be essential for catalysis and substrate binding as both belong to the same class of enzymes. The DNA repair (5'-3' exonuclease) and proof-reading (3'-5' exonuclease) functions did not show any super imposable regions. However, interestingly the C terminal region (polymerase domain) shows few super imposable regions including the catalytic K, YG pair and the invariant R (marked 1-9). Additionally, a GT, YM diad and an LPL/V triad are also observed, in addition to few single amino acid conservations. The catalytic, template and substrate binding motifs are highlighted. The YG gate keeper motif and the catalytic K are strictly conserved (including distance conservation) in both DNA dependent DNA polymerases and DNA dependent RNA polymerases. This suggests that the polymerase reaction is accomplished using the same set of amino acids in DNA and RNA polymerases.

CLUSTAL O(1.2.4) MSA of *E. coli* DNA polymerases I, T7, T3, SP6 and K11 RNA polymerases



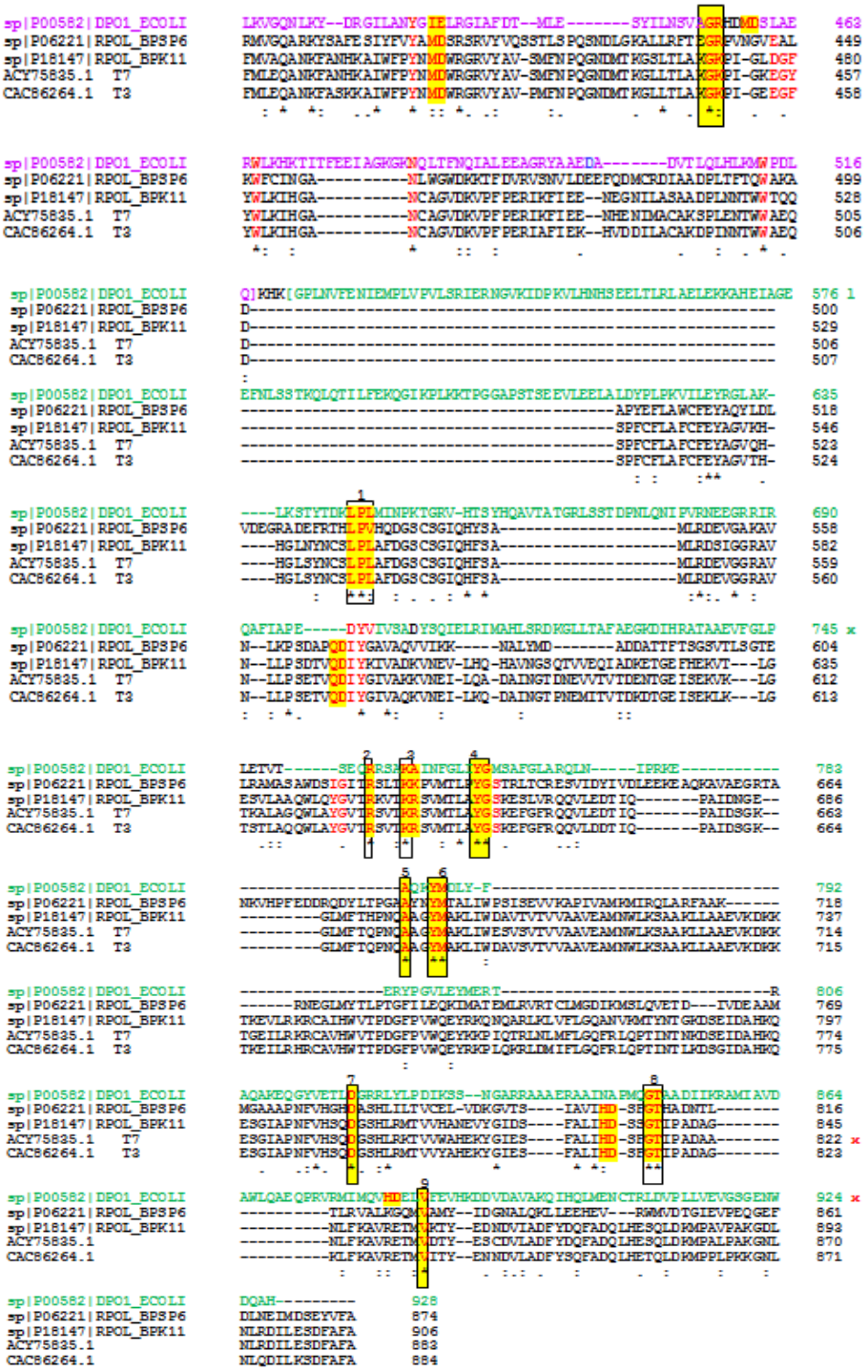


Fig. 7. Multiple sequence alignment *E. coli* DNA pol I, T3, T7, K11 and SP6 RNA polymerases
sp|P00582|DPO1_ECOLI, adjust spacing *Escherichia coli*
sp|P06221|RPOL_BPSP6, Enterobacteria phage SP6
sp|P18147|RPOL_BPK11, Enterobacteria phage K11
ACY75835.1, Enterobacteria phage T7
CAC86264.1, Enterobacteria phage T3

(E. coli DNA polymerase I is made up of three domains, viz. amino acids 1-323 constitute the 5'-3' exonuclease domain or DNA repair domain (323 amino acids length); amino acids 324-517 constitute the 5'-3' exonuclease domain or proof-reading domain (194 amino acids length); and amino acids 521-928 constitute the polymerase domain (408 amino acids length). The proof-reading and the polymerase domain from 324 to 928 amino acids (605 amino acids length) is known as Klenow polymerase. All three domains are shown in different colours).

3.2 Dissection of DNA and RNA Polymerases

Different domains of the *E. coli* DNA polymerase I are shown in Fig. 8 [4]. *E. coli* DNA polymerase I is made up of three domains, viz. amino acids 1-323 constitute the 5'-3' exonuclease domain or DNA repair domain (323 amino acids length); amino acids 324-517 constitute the 5'-3' exonuclease domain or proof reading domain (194 amino acids in length); and amino acids 521-928 constitute the polymerase domain (408 amino acids in length). The proof reading and the polymerase domain from 324 to 928 amino acids (605 amino acids in length) is also known as Klenow polymerase. In T7 RNA polymerase also the polymerase domain is found in the C terminal region starting from 507-883 (376 amino acids length (Fig. 8). In T7 RNA polymerase, the amino-terminal region is reported to be involved in promoter recognition and DNA melting functions [7].

3.3 Analysis of Polymerase Active Site in the RNA and DNA Polymerases

Both the RNA and DNA polymerases belong to the same Main class (Transferases) and come under the sub class nucleotidyl transferases (EC 2.7.7.6 and EC 2.7.7.7). Therefore, both the enzymes might be of similar structure and use similar mechanism of action. It is well established by biochemical, genetic and site-directed mutagenesis that the polymerase I active is at the carboxy terminal domain (CTD) of the enzyme (Fig. 8). The CTD contains KA, a YG pair and an invariant R at -4 where the KA pair involves in catalysis and the YG pair and R act as steric gate allowing only the dNTPs for polymerization. It is interesting to note that the polymerase active is in CTD of T7 RNA polymerase [8] and similar conserved amino acids are also found in all the SSU DNA dependent RNA polymerases studied (Figs. 2-7). Another interesting finding is that in the viral polymerases an additional YG is found in the downstream exactly at the same distance but downstream.

3.4 Distance Conservation between Catalytic K and YG Pair in DNA and RNA Polymerases

3.4.1 Catalytic K and YG pair in DNA polymerases

It is interesting to note that the catalytic amino acid K and the gate keeper pair YG are completely conserved in different polymerases from a diverse group of organisms (Table1). The mismatched regions in some of the polymerases were aligned as suggested by Palanivelu [9]. (In this analysis only the amino acids around the active site regions of different DNA polymerases from different sources ranging from virus to plant and animals were selected and analyzed by T-COFFEE advanced version). Table 1 summarizes the above findings. It is interesting to note that irrespective of the type of polymerases and their origin, all of them showed a completely conserved K at the catalytic site and YG pair at the steric gate position. A distance conservation is also observed in all these polymerases, (i.e.), the YG pair is 8/9 amino acids downstream of the catalytic K. The invariant G is found in all these DNA and RNA polymerases, probably as it is the only amino acid, achiral and fit into both hydrophilic and hydrophobic environments.

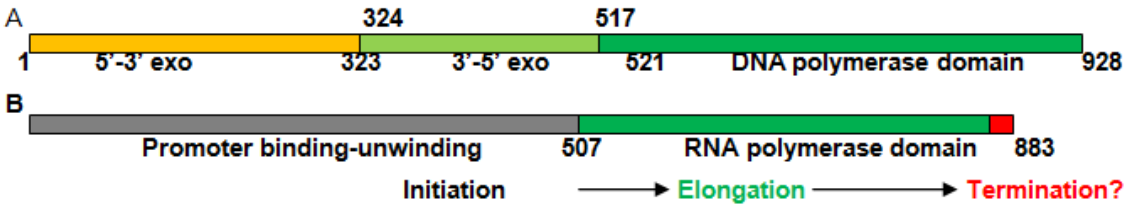


Fig. 8. Dissection of *E. coli* DNA polymerase I (A) and T7 RNA polymerase (B)

3.4.2 Catalytic K and YG pair in RNA polymerases

The distance between the catalytic K and YG pair is remarkably conserved among both the SSU RNA polymerases from viruses, mitochondria and chloroplasts and the DNA polymerases [4]. From the catalytic amino acid K, the YG pair is exactly the 8th amino acid upstream in all the SSU RNA polymerases from viruses, mitochondria and chloroplasts (Table 2). It is interesting to note that in the viral polymerases, an additional YG pair is placed exactly at the same distance but towards downstream (-8) of the catalytic K. Thus, the viral polymerases show two YG pairs placed exactly at the same distances from the catalytic amino acid on both the sides. Such additional YG pair is not found in the mitochondrial and chloroplast RNA polymerases and also in the *E. coli* DNA polymerase I. This suggests that the YG pair may bind on both the strands and placing the catalytic K in the middle and all three moving downstream incorporating the NTPs. In fact. Kotsyuk et al. [5] have shown that the T7 polymerase requires both the strands for activity and there was no activity when single-stranded DNA was used as the substrate.

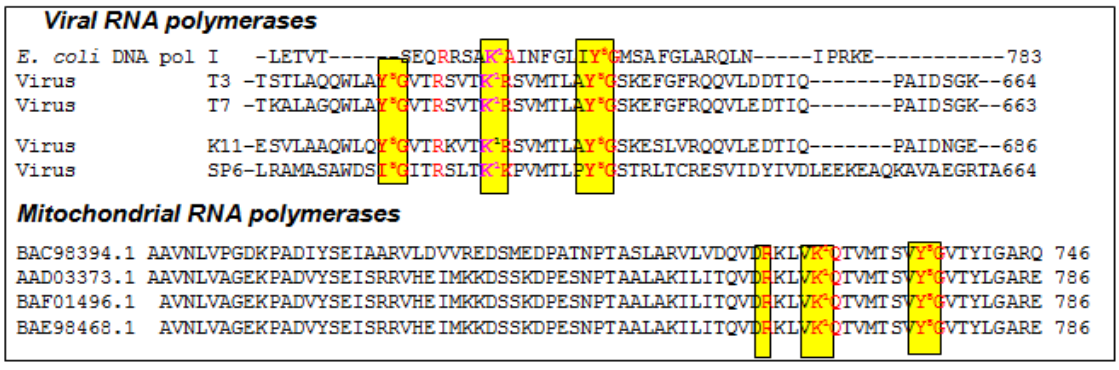
It is interesting to note that the YG pair appears to be specific for polymerases using DNA as the template (including the prokaryotic and eukaryotic multi-subunit RNA polymerases, data not shown) as it is not reported in RNA dependent polymerases where they use RNA as the template [2].

Table 1. The catalytic amino acid (K) and gate keeper pair (YG) in different polymerases from diverse sources

T4 DNA pol	546 ATLANTNQLNR**K**[1]**ILINSLY**[8]**G**ALGNIH
Human HSV 1	800 AVLLDKQQAAI**K**[1]**VVCNSVY**[8]**G**FTGVQH
E. coli DNA pol I	748 TVTSEQRRSA**K**[1]**AINFGLIY**[9]**G**MSAFGLAR
E. coli DNA pol II	482 RQGNKPLSQAL**K**[1]**IIMNAFY**[8]**G**VLGTTA
E. coli DNA pol III (alpha subunit)	663 YPDVQWQHESL**K**[1]**PVLEPTY**[8]**G**IILYQE
P. furiosus DNA pol	477 KILLDYRQKAI**K**[1]**LLANSFY**[8]**G**YYGYAK
Yeast alpha DNA pol	933 RVQCDIRQQAL**K**[1]**LTANSMY**[8]**G**CLGYVN
Human alpha DNA pol	939 ILQYDIRQKAL**K**[1]**LTANSMY**[8]**G**CLGFSY
Human Gamma DNA pol	917 TTVGISREHA**K**[1]**IFNYGRIY**[9]**G**AGQPFAER
Human Delta DNA pol (Catalytic subunit)	683 RQVLDGRQLAL**K**[1]**VSANSVY**[8]**G**FTGAQV
Human epsilon DNA pol	798 EVLYDSLQLAH**K**[1]**CILNSFY**[8]**G**YVMRKGAR
A. thaliana Delta DNA pol (Catalytic subunit)	679 KAVLDGRQLAL**K**[1]**ISANSVY**[8]**G**FTGATV

N.B: Some of the above polymerases did not align in T COFFEE advanced version. So the conserved regions were selected and aligned as suggested by Palanivelu [9].Table 1 from Palanivelu [4]).

Table 2. The catalytic amino acid (K) and gate keeper pair (YG) in different SSU RNA polymerases from diverse sources



3.4.3 Three critical pairs of amino acids in DNA and RNA polymerases and their possible function(s)

These analyses have revealed three critical pairs of amino acids in all these polymerases belonging to different types and origins. Table 3 shows these three critical pairs from different representative DNA and RNA polymerases.

The pair 1 involves in polymerization, (the Lys functions as proton abstractor); and the Arg with its guanidinium group make multiple hydrogen bonds to the NTP/dNTP.

The pair 2 (YG pair) acts as "*steric gate"* and involves in template binding and allows only NTPs in RNA polymerases (dNTPs in DNA polymerases) at polymerization site possibly with other conserved amino acids and

The pair 3 involves as "*charge shielder"* of NTPs/dNTPs through a Mg ion (Table 3) and orients the α-phosphates of NTPs/dNTPs for polymerization.

3.4.4 Similarities in the active sites of DNA and RNA polymerases

3.4.4.1 The invariant K and its role in E. coli DNA polymerase I and T7 RNA polymerase

The DNA and RNA polymerase active sites were probed by a variety of techniques:

In the *E. coli* DNA polymerases I, DNase foot-printing assay using DNase I and methidium-propyl EDTA-Fe^{2+} indicated that the enzyme binds to the primer terminus and covers 8 base pairs [10].

Photo affinity labeling of the enzyme with dNTP analogue, 8-azido-dATP, and sequencing of the labeled peptide, identified Tyr^{766} at the active site of the enzyme [11]. Thus, the foot printing and photo cross-linking experiment has suggested the Tyr^{766} in the active site.

Table 3. Critical pairs of amino acids found in DNA polymerases and RNA polymerases

Enzyme	Arg/Lys	Tyr//Gly	Asp//Asp
DNA polymerases			
T_7 pol	Arg518/Lys522	Tyr530/Gly531	Asp475/Asp654
E. coli pol I	Arg754/Lys758	Tyr766/Gly 767	Asp705/Asp882
E. coli pol II	Leu*523/Arg527	Phe533/Gly534	Asp452/ Asp545
E. coli pol III	^----/Lys758	Tyr764/^Ala765	Asp405/Asp733
Human α	Lys947/Lys950	Tyr957/Gly958	Asp860/Asp1004
Human β	Leu*163/Lys168	Tyr 173/Leu*174	Asp192/Asp256
Human γ	Arg943/Lys947	Tyr955/Gly956	Asp890/Asp935
Human δ	Arg689/Lys694	Tyr701/Gly702	Asp602/Asp757
Yeast ε	Leu*819/Lys824	Tyr831/Gly832	Asp669/ Asp2118
Viral RNA polymerases			
T7	Arg627/Lys631	Tyr639/Gly640	Asp537/Asp812
T3	Arg628/Lys632	Tyr640/Gly641	Asp538/Asp813
Chloroplast RNA polymerases[#]			
A Thaliana	Arg725/Lys729	Tyr737/Gly738	Asp654/Asp886
O. Sativa	Arg765/Lys769	Tyr777/Gly778	Asp694/Asp926
Mitochondrial RNA polymerases[#]			
A Thaliana	Arg748/Lys752	Tyr760/Gly761	Asp677/Asp909
N. sylvestris	Arg774/Lys778	Tyr786/Gly787	Asp703/Asp935

Based on multiple sequence analysis

** Instead of Arg, a Leu is found at the corresponding position in the repair polymerases, viz., pol II and pol β.*

^ No Arg or Leu is found near vicinity of the probable catalytic K. A good number of prokaryotic replicative polymerases (pol III) had an Ala adjacent to the Tyr)

In almost all the pol IV polymerases, only a G (PXG) is seen at the 11th position from the catalytic K; no regular gate keeper Y is found which possibly explains the error-prone nature of these polymerases

The ε polymerases also maintain a Leu near the catalytic K, as it is also involved in DNA repair. Pol ε's main function is to extend the leading strand during replication while Pol δ is involved in the lagging strand synthesis. The most striking difference between the two DNA polymerases is that processive DNA synthesis by DNA polymerase delta is dependent on proliferating cell nuclear antigen (PCNA), a replication factor, while DNA polymerase epsilon is inherently processive.

[#] RNA polymerase data based on MSA delete single bracket

However, Basu and Modak [12], who have probed the polymerase active site with pyridoxal phosphate, found Lys^{758} at the active site (pyridoxal phosphate binds competitively to the dNTP site through Schiff's base formation and covalently links the amino acid involved possibly in polymerization reaction). These results suggest that the polymerase active site is in the bigger domain remote from the 3'→5' exonuclease activity and totally not connected to the dNMP site. Similar observations were made with an adenovirus DNA polymerase, e.g., the pyridoxal phosphate modification of an adenovirus DNA polymerase resulted in the loss of DNA polymerase activity, whereas the 3'- 5' exonuclease activity was unaffected. Inhibition of adenovirus DNA polymerase by pyridoxal phosphate was time-dependent and displayed saturation kinetics [13]. It is interesting to note that Zaldivar et al. [14] have shown that not only in DNA polymerases but also in RNA polymerases I and II of rat liver and RNA polymerase I of yeast, were also inactivated by pyridoxal phosphate and hence suggested a possible involvement of a Lys residue in the catalytic site of RNA polymerases too.

Thus, both the Lys and Tyr are completely conserved in DNA polymerases analyzed by Palanivelu [4] and RNA polymerases (this communication). The phi 29 viral DNA polymerase shares several regions of amino acid similarity with other alpha-like DNA polymerases. Among them, the conserved region characterized by the amino acid motif "**K**x3NSx**YG**" has been proposed to form part of the polymerization active site of alpha-like DNA polymerases [15]. However, by MSA analysis, these polymerases have shown a completely conserved R exactly at the 4th position downstream from the catalytic K and hence should be also included the template binding and catalysis and thus active site motif in both DNA and RNA polymerases is "$\mathbf{R}^{-4}$xxx$\mathbf{K}^{1}$xxxxxx$\mathbf{Y}^{+8}\mathbf{G}$". By using a library with totally random nucleotides at five different codons (R659, R660, K663, F667, and G668), Suzuki et

al. [16] confirmed that R^{659} and K^{663} were immutable in the DNA polymerase from *T. aquaticus* (R^{754} and K^{758} in *E. coli* DNA polymerase I, respectively),

The following observations also support As the Lys is completely conserved in both the types of polymerases [4 and this communication]) *it is proposed that the catalytic amino acid could be the completely conserved K in both the RNA and DNA polymerases*

Furthermore, Lys is the active site amino acid in NAD- and ATP dependent ligases, and also GTP dependent mRNA capping enzymes, which are all involved in making a phosphodiester bond as in polymerases [17].

Like DNA polymerases the DNA ligases are also inhibited by pyridoxal 5'-phosphate indicating the presence of a K at the catalytic domain of the enzyme [18]. Both the types of ligases (ATP-dependent and NAD-dependent DNA ligases) from various organisms showed a highly conserved motif KY/I/VDGXR with the reactive K residue, followed by a Y or a hydrophobic amino acid [17].

Interestingly, not only in DNA ligases, but also in RNA ligases the catalytic K is conserved [18].

In *E. coli* DNA polymerase I Y^{766} and Y^{758} are found to be in close proximity to the 3'-OH of the primer and interestingly, such proximity is completely conserved in both the types of polymerases. Further analysis by site-directed mutagenesis, Doublie and Ellenberger [19] and Astatke et al. [20] have shown that the critical Y may possibly be involved in template recognition and dNTP selection in DNA polymerases [21] and the same function is proposed for RNA polymerases as well in this communication. It is interesting to note that a highly conserved Tyr residue in reverse transcriptase controls substrate selection. It is interesting to note that the highly conserve Y^{955} residue is critical for nucleotide recognition among Family A DNA polymerases, i.e., γ polymerases from eukaryotes. Furthermore, Y^{955} is a highly conserved residue among a wide variety of DNA polymerases (Table 1). Further proof of Y^{766} involvement in nucleotide selection was obtained from site-directed mutagenesis of Y^{766}; substitution of an equivalent amino acid as in $Y^{766}\rightarrow$F substitution in the Klenow polymerase did not show an appreciable increase in nucleotide misinsertion; however, substitution with Ala or Ser generated an error-prone DNA polymerase attributable to decreased stringency for selection of dNTPs [19]. Interestingly the YG doublet is highly conserved and found to be a common pair in different types of DNA and RNA polymerases (Table 4).

Further proof is provided by crystallographic analysis of T7 DNA polymerase. The T7 DNA replication complex at 2.2 Å resolution have shown that the invariant K^{522} (≡ K^{758} in *E. coli* DNA pol I) actually makes contact with the α-phosphate of dNTP [22].

Since, the mechanism of action for polymerization reactions of RNA polymerases, proposed in this article, is based on a proton abstraction at the catalytic site amino acid, K is placed as the catalytic amino acid. The other active site amino acids, viz. the YG pair and possibly with other conserved amino acid(s), holds the complementary base inserted by the finger domain onto the catalytic site, the catalytic K adds the NTP to the 3'-OH. The reaction essentially occurs through proton abstraction by K followed by an electrophilic-nucleophilic attack at the growing 3" end (Figs. 9.1-9.4).

3.4.4.2 T7 polymerase used as the model enzyme for studying transcription

Perhaps the most widely studied single-subunit RNA polymerase is bacteriophage similar to the *E. coli* DNA polymerase I for polymerase family. The common feature of all these SSU RNA polymerases from T7, T3, SP6, and K11, mitochondrial and chloroplast, is their simpler structure compared to prokaryotic and eukaryotic multi-subunit RNA polymerases which are more complex. Interestingly, even though they are single-subunit RNA polymerases, they are able to perform the complete transcriptional cycle in the absence of additional protein factors. The single-subunit composition, relatively low molecular weight, makes the T7 RNA polymerases the most convenient model for investigating the physicochemical aspects of transcription and its catalytic mechanism. Furthermore, the enzyme can be produced in large amounts for structural analysis.

3.4.4.3 Properties of T7 RNA polymerase

T7 RNA polymerase was first isolated from T7-infected *E. coli* cells in 1970 [23]. It has 883 amino acids with a molecular mass of 98,092 Daltons, optimally active in the pH range 8.0-9.0 and the elongation rate is 100-200 nucleotides/sec. The T7 polymerase requires a

double-stranded DNA template and 5- 10 mM Mg^{2+} as a cofactor for the optimal synthesis of RNA [2]. It is an extremely promoter-specific enzyme and transcribes only DNA downstream of a T7 promoter (TAATACGACTCACTATAG) and the transcription begins with the 3' G. It has a very low error rate. Interestingly, the 3D structures of polymerization domains are very similar in DNA and RNA polymerases, including the T7 RNA polymerase [8]; they all resemble a right hand and the sub domains are referred to as "palm", "thumb", and "fingers". However, the T7 family of RNA polymerases is structurally and evolutionarily distinct from the multi-subunit family of RNA polymerases of bacterial and eukaryotic families and is not inhibited by the antibiotic, rifampicin. In biotechnology applications, T7 RNA polymerase is used to transcribe DNA in many modern-day vectors that have been cloned into such vectors.

3.4.4.4 Analysis of active site and metal binding site(s) of T7 RNA polymerase

Table 4 shows the summary of site-directed mutagenesis of the T7 RNA polymerase active site [24]. The catalytic K^{631} when modified with either G or L or R, only partially inactivated the T7 RNA polymerase whereas in DNA polymerase I, it was completely inactivated. However, in another site-directed mutagenesis experiment, Osumi-Davis et al. [25] have shown when the K^{631} is modified to M, T7 RNA polymerase has lost almost all the activity. Interestingly, the Y^{639} of the YG pair is also essential for its activity as the modification of this critical Y^{639} yielded no activity as expected. Site-directed mutagenesis experiments have also shown other amino acids like P^{563}, Y^{571}, T^{636}, F^{646} are also important for the activity of the enzyme (Table 4). By electron paramagnetic resonance spectroscopy, flow-dialysis and transcription analysis, the D537 and D812 in bacteriophage T7 RNA polymerase are found to be as metal ion-binding sites and are essential in the catalytic mechanism [26, 27].

3.4.5 Mechanism of NTP and dNTP discrimination in SSU RNA polymerases

Though the RNA polymerases, DNA polymerases or reverse transcriptases are divergent, the overall 3 D structures are found to be very similar and follow right handed palm, fingers and thump shape. The RNA polymerases, DNA polymerases use the same catalytic amino acid and gate keeper pair and an invariant R in their catalytic motif. Then the most intriguing question is how the RNA polymerases discriminate NTPs from dNTPs and allow only NTPs to the polymerization site. This problem was solved by an interesting mutagenesis experiment by Kotsyuk et al. [5]. They observed that the YG pair in viral RNA polymerases is characterized by a unique distribution of invariant hydroxyl-containing amino acids like S and T, whereas no such

Table 4. Summary of site-directed mutagenesis results on T7 RNA polymerase

Amino acid position and modification	Result	Reference
$K^{172} \longrightarrow$ L	No change in activity	[28]
$P^{563} \longrightarrow$ A	Inactivated	[24]
$Y^{571} \longrightarrow$ S	Inactivated	ibid
$K^{631} \longrightarrow$ G/L/R	Partially inactivated	ibid
$T^{636} \longrightarrow$ P	Inactivated	ibid
$Y^{639} \longrightarrow$ D	Inactivated	ibid
$F^{646} \longrightarrow$ C	Inactivated	ibid
$D^{537} \longrightarrow$ N	Inactivated (Total)	[25]
$D^{812} \longrightarrow$ N	Inactivated (Total)	ibid
$K^{631} \longrightarrow$ M	Inactivated (1% activity)	[29]
$Y^{639} \longrightarrow$ F	Fully active*	ibid
$S^{641} \longrightarrow$ A	No RNAP but shows DNAP	[5]

**kinetic parameters are somewhat different*
 # RNAP/DNAP = RNA polymerases/ DNA polymerase activity

regularity is seen in DNA polymerases [4]. In order to find out whether this unique Ser^{641} in T7 RNA polymerase ($Y^{639}GS^{641}$) play any role in the discrimination between NTPs and dNTPs, they made a single amino acid substitution and the S^{641} was modified to Ala. The mutant enzyme was purified to homogeneity and found to their surprise that the mutant enzyme allowed dNTPs also and the T7RNA polymerase lost RNA polymerase activity and exhibited DNA polymerase activity. The Ser hydroxyl likely recognizes the 2'-OH in the NTPs and possibly makes a hydrogen bond and discriminate dNTPs, where they lack a 2'-OH. If you have close look at the other SSU RNA polymerase from mitochondria and chloroplasts a functionally equivalent T is placed in the vicinity of the YG pair as YGxT. A similar observation was made by Cermakian et al. [3] The invariant T in these polymerases possibly involves in the NTP and dNTP discrimination as in the case of T7 RNA polymerases.

3.4.6 Mechanism of action of T7 RNA polymerase

The mechanism of action of T7 RNA polymerase is proposed based on the data obtained by MSA and data already available by biochemical, site-directed mutagenesis and X-ray crystallographic analysis. X-ray crystallographic analysis of T7 polymerase have shown that the B motif is located in the 'finger' subdomain, close to motifs A and C with both these motifs likely to form the active site. The side chain radicals of the three invariant amino acids (R^{627}, K^{631} and Y^{639}) are found to be directed towards the substrate binding cleft [30]. Temiakov et al. [31] have also shown that Y^{639} is mainly involved in discrimination of ribose versus deoxyribose substrates and the substrate selection occurs prior to the isomerization to the catalytically active conformation. However, an invariant Y is also found in DNA polymerases as well, at the same distance from the catalytic K [4]. Therefore, the presence of an invariant S or T adjacent to YG pair in these RNA polymerases was found to be playing an important role in substrate selection [5,3]. Whitney Yin and Steitz [32] have observed two divalent metal ions in the active site of T7 RNA polymerase; metal ion A is associated exclusively with the 3′ end of RNA in the product complex while metal ion B remains bound to the product pyrophosphate as well as the catalytic carboxylate.

The proof-reading mechanism is well established in DNA polymerases [4]. However, it is poorly understood in RNA polymerases. Maintaining high fidelity during transcription is essential for the accurate transfer of genetic information from DNA to RNA. (RNA polymerases generally misincorporate only one wrong nucleotide/~100000 bases). As RNA polymerases are also Zn metalloenzymes (possibly the metal ion A, which is associated exclusively with the 3′ end of RNA as discussed elsewhere) the Zn-mediated deletion [4] of the misincorporated NTP could be a possible mechanism, as the enzyme stalls at every misincorporation like DNA polymerases. Zn mediated hydrolysis could be also the possible mechanism for RNA cleavage followed by dissociation at transcription termination, where the RNA polymerase again stalls at the termination site.

Fig. 9.1. Watson-Crick base pairing of the incoming nucleotide with the template and nucleotide discrimination by steric gate amino acids Tyr, Gly and Ser

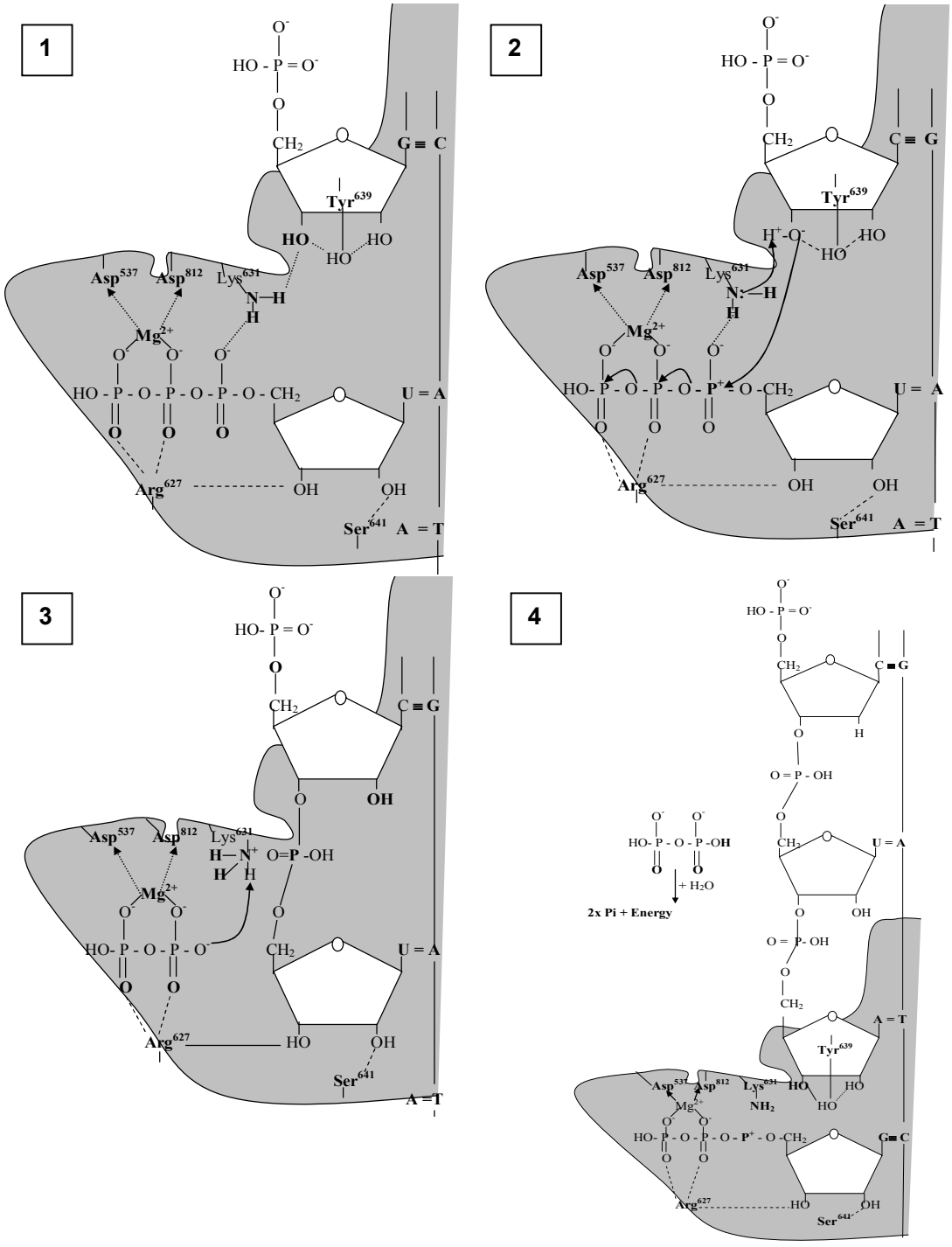
Fig. 9.2. Electronic transition at the active site for proton abstraction and an electrophilic and nucleophilic attack

Fig.9.3. Proton abstraction by the active site amino acid Lys with simultaneous formation of 3'→5' phosphodiester bond with the incoming NTPs

Fig. 9.4. Transfer of the proton from Lys to pyrophosphate and formation of inorganic pyrophosphate and the translocation of the enzyme to next complementary nucleotide in-position that is to be polymerized.

3.4.7 Other conserved amino acids and regions in these polymerases

It should also be noted that the above discussed conserved motifs and amino acids form only the substrate binding and catalytic cores. Apart from these, there are a large number of single amino acid invariants (Ys, Ws, Cs, Ps and Gs,) diads, triads and long conserved stretches of amino acids in all these polymerases (Figs. 2–7). A good number of highly conserved Ps in these polymerases is implicated in making the necessary bents on the enzyme's structure during substrate binding, polymerization and translocation processes. The long conserved stretches of amino acids might be required to



Figs. 9.1-9.4. Steps proposed in the polymerization reaction of T7 RNA polymerase

make the correct, unique 3D structures. The highly conserved Cs might be useful for making the disulphide bridges which make the enzyme more compact and stable.

4. CONCLUSION

MSA have shown that a basic amino acid K, a YG pair and an invariant R and S/T are highly conserved in all SSU RNA polymerases.

Distance conservation is also found among the conserved motifs and amino acids among these RNA polymerases. Site-directed mutagenesis, biochemical and X-ray crystallographic analyses of T7 RNA polymerase have also suggested their involvement in substrate binding and catalysis. Based on these results, a plausible mechanism of action is proposed for the polymerization reactions for T7 RNA polymerase as the model enzyme.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. S. Krishnaswamy, Centre of Excellence in Bioinformatics (Retd.) and Dr. R. Usha, Department of Plant Biotechnology (Retd.), School of Biotechnology, Madurai Kamaraj University, Madurai, for their useful suggestions on the manuscript.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Hurwitz J. The discovery of RNA polymerase. J Biol Chem. 2005;280: 42477–85.
2. Tunitskaya VL, Kochetkov SN. Structural and functional analysis of bacteriophage T7 RNA polymerase. Biochemistry (Moscow). 2002;67:1124–1135.
3. Cermakian N, Ikeda TM, Miramontes P, Lang BF, Gray MW, Cedergren R. On the evolution of the single-subunit RNA polymerases. J Mol Evol. 1997;45:671–681.
4. Palanivelu P. DNA polymerases – An insight into their active sites and mechanism of action, Int. J. Biochem. Res. & Rev. 2013;3:205-247.
5. Kostyuk SM, Dragan DL, Lyakhov VO, Rechinsky VL, Tunitskaya BK. Chernov SN. Kochetkov E. Mutants ofT7 RNA polymerase that are able to synthesize both RNA and DNA. FEBS Letters. 1995;369:165-168.
6. Konig J, Baumann S, Koepke J, Pohlmann T, Zarnack K, Feldbrugge M. The fungal RNA-binding protein Rrm4 mediates long-distance transport of ubi1 and rho3 mRNAs. EMBO J. 2009;28:1855-1866.
7. Cheetham GM, Jeruzalmi D, Steitz TA, Structural basis for initiation of transcription from an RNA polymerase-promoter complex. Nature. 1999;399:80-83.
8. Sousa R, Je Chung Y, Rose JP, Wang BC. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. Nature. 1993;364:593-599.
9. Palanivelu P. Multiple sequence analysis of polygalacturonases and invertases and phase shift in conserved motifs. Indian J Biotechnol. 2007;6:24-30.
10. Brenowitz M, Senear DF, Shea MA Ackers GK. Quantitative DNase footprint titration: A method for studying protein-DNA interactions. Methods Enzymol. 1986;130: 132-181.
11. Rush J, Konisberg W H. Photoaffinity labeling of the Klenow fragment with 8-azido-dATP. J Biol Chem. 1990;265:4821-27.
12. Basu A, Modak MJ. Affinity labeling of *E. coli* DNA polymerase I by pyridoxal 5'-phosphate. Biochemistry. 1987;26:704-1709.
13. Monaghan N, Hay RT. Pyridoxal 5'-phosphate inhibition of adenovirus DNA polymerase. J Biol Chem. 1996;271: 24242-8.
14. Zaldivar MJ, Bull P, Venegas A, Valenzuela P Inactivation of rat liver RNA polymerases I and II and yeast RNA polymerase I by pyrodixal 5'-phosphate. Evidence for the participation of lysyl residues at the active site. Biochemistry. 1975;14:4907-11.
15. Blasco MA, Lázaro JM, Blanco L, Salas M. Phi 29 DNA polymerase active site. The conserved amino acid motif "Kx3NSxYG" is involved in template-primer binding and dNTP selection, Proc Natl Acad Sci. (USA). 1997;94:1619–1622.
16. Suzuki M, Baskint D, Hoodt L, Loeb, LA. Random mutagenesis of *Thermus aquaticus* DNA polymerase I: Concordance of immutable sites in vivo with the crystal structure. Proc Natl Acad Sci. (USA). 1996;93:9670-9675.
17. Doublie S, Tabor S, Long AM, Richardson, CC, Ellengerger T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. Nature. 1998;391: 251-258.
18. Sriskanda V, Schwer B, Ho K, Shuman S. Mutational analysis of *Escherichia coli* DNA ligase identities amino acids required for nick-ligation *in vitro* and for *in vivo* complementation of their growth of yeast

cells for CDC9 and LIG4, Nucleic acids Res. 1999;27:3953-3963.
19. Tomkinson AE, Totty NF, Ginsburg M, Lindahl T. Location of the active site for enzyme-adenylate formation in DNA ligases, Pro Natl Acad Sci. (USA). 1991;88:400-404.
20. Astatke M, Grindley N D, Joyce C M. How *E. coli* DNA polymerase I (Klenow fragment) distinguishes between deoxy- and dideoxynucleotides, J Mol Biol. 1998;278:147-165.
21. Kaushik N, Pandey VN, Modak MJ. Significance of the O-helix residues of Escherichia coli DNA polymerase I in DNA synthesis; dynamics of the dNTP binding pocket. Biochemistry. 1996;35:7256-66.
22. Derbyshire V, Freemont PS, Sanderson MR, Beese L, Friedman JM, Joyce CM , Steitz TA. Genetic and crystallographic studies of the 3,5'-exonucleolyitc site of DNA polymerase I, Science. 1988;240: 199-201.
23. Chamberline M, McGrath J, Wakshell L. New RNA polymerase from *Escherichia coli* infected with Bacteriophage T7. Nature. 1970;228:227-231.
24. Rechinsky VO, Kostyuk DA, Lyakhov DL, Chernov BK, Kochetkov SN. Random mntagenesis of the gene for bacteriophage T7 RNA polymerase. Mol Gen Gen. 1993;238:455–458.
25. Osumi-Davis P de Aguilera MC, Woody RW, Woody AY. Asp537, Asp812 are essential and Lys631, His811 are catalytically significant in bacteriophage T7 RNA polymerase activity. J Mol Biol. 1992;226:37-45.
26. Osumi-Davis P, Sreerama N, Volkin DB, Middaugh CR, Woody RW, Woody AYM. Bacteriophage T7 RNA polymerase and its active-site mutants. Kinetic, spectroscopic and calorimetric characterization. J Mol Biol. 1994;237:5-19.
27. Woody AY, Eaton SS, Osumi-Davis PA, Woody RW, Asp537 and Asp812 in bacteriophage T7 RNA polymerase as metal ion-binding sites studied by EPR, flow-dialysis, and transcription. Biochemistry. 1996;35:144-52.
28. Lyakhov DL, Ilgenfrtiz H, Chernov BK, Dragan SM, Rechinsky VO, Pokholok DK, Tunitskaya VL, Kochetkov SN. Site-specific mutagenesis of residue Lys-172 of phage T7 RNA polymerase: Characterization of transcription properties of mutant proteins. Mol Biol. (Moscow). 1992;26:1022-1035.
29. Maksimova TG, Mustayev AA, Zaychikov EF, Lyakhov DL, Tunitskaya VL, Akbarov AK, Luchin SV, Rechinsky VO, Chernov BK, Kochetkov SN. Lys631 residue in the active site of the bacteriophage T7 RNA polymerase. Affinity labeling and site-directed mutagenesis, Eur J Biochem. 1991;195:841-847.
30. Doublie S, Ellenberger T. The mechanism of action of T7 DNA polymerase. Curr Opin Struct Biol. 1998;8:704-712.
31. Temiakov D, Patlan V, Anikin M, McAllister WT, Yokoyama S, Vassylyev DG**.** Structural basis for substrate selection by T7 RNA polymerase. Cell. 2004;116:381-391.
32. Whitney Yin Y, Steitz TA. The Structural Mechanism of Translocation and Helicase Activity in T7 RNA Polymerase. Cell. 2004;116: 393-404.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history/22801