



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

P. Palanivelu^{1*}

¹Department 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,

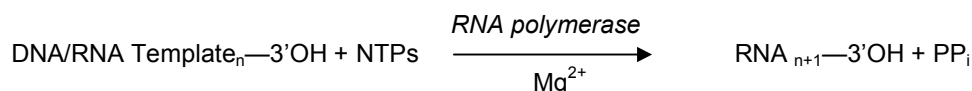


Fig. 1a. Dynamics of RNA polymerization

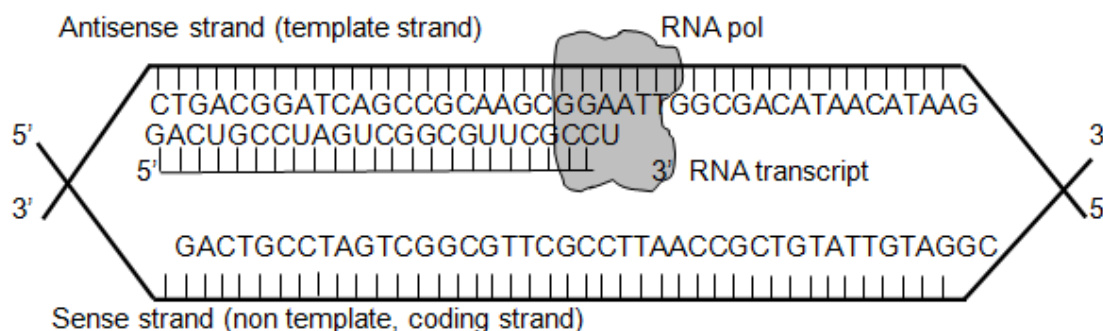


Fig. 1b. Process of RNA transcription using the template strand

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 polycistronic.

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          -----MQDLHAIQLQLEEFMFNGGIRRFEDQQRQI  31
AAZ72968.1                    -MSVISIDKHFSDVSNAIPEFNLLADHYGQDLAVKQLQLEHEAYTEGERRFIKNLERQT  59
YP_009044255.1                -MSVISIDKHFSDVSNAIPEFNLLADHYGQDLAVKQLQLEHEAYTEGERRFIKNLERQT  59
AEH41021.1                    -MNALNIARNDFSEIELAAIPYNYLSEHYGDRLLAREQLALEHEAYELGEQRFLKMLERQV  59
sp|P18147|RPOL_BPK11         -MNALNIGRNDFSEIELAAIPYNYLSEHYGDRLLAREQLALEHEAYELGRQRFLKMLERQV  59
ACY75835.1                    -MNTINIAKNDFSDIELAAPFNLLADHYGERLAREQLALEHESYEMGEARFRKMFERQL  59
ACOS7213.1                    -MNTINIAKNDFSDIELAAPFNLLADHYGERLAREQLALEHESYEMGEARFRKMFERQL  59
tr|C62CU5|C62CU5_LAMBD       -MNTINIAKNDFSDIELAAPFNLLADHYGERLAREQLALEHESYEMGEARFRKMFERQL  59
CAC86264.1                    MNI IENI EKND FSEI ELAAI PFNTLADHYG SALAKEQLALEHESYELGER RFLKMLERQA  60
                                     .          ** **.* : * **      : **

```

```
sp|P06221|RPOL_BPSP6      AAGSESDTAWNRRLLSELIAPMAEGIQAYKEEYEGKKGRAPRALAFLQC----- 80
AAZ72968.1                 ERGELADNQAQVAKPLMQTLVFKIQAQAVKEHHEGPDGKLGKLDSTRPSVAFTML-----STEE 112
YP_009044255.1            ERGELADNQAQVAKPLMQTLVFKIQAQAVKEHHEGPDGKLGKLDSTRPSVAFTML-----STEE 112
AEH41021.1                 KAGEFADNVAAPKPLVLTLLHPQLTKRIDDWKEEQANARGKKPRAYYPPIKHGVASELALSMG 119
sp|P18147|RPOL BPK11      KAGEFADNVAAPKPLVLTLLHPQLTKRIDDWKEEQANARGKKPRAYYPPIKHGVASELALSMG 119
ACY75835.1                 KAGEVADNAAAKPLITLLPKMIARINDWFEVVKAKRGKRPATFQFLQE----- 108
AC057213.1                 KAGEVADNAAAKPLITLLPKMIARINDWFEVVKAKRGKRPATFQFLQE----- 108
tr|C6ZCU5|C6ZCU5_LAMB    KAGEVADNAAAKPLITLLPKMIARINDWFEVVKAKRGKRPATFQFLQE----- 108
CAC86264.1                 KAGEIADNAAAKPLLATLLPKLTTRIVEWLEEYASKKGRKPSAYAPLQL----- 109
* . : * . : * : * : : : * . :
sp|P06221|RPOL_BPSP6      -----VENEVAAYITMKVMDMLNT--DATLQAIAMSVAEIERIDQVRFSKLEGHAA 129
AAZ72968.1                 RAVKDRSLRISCESAAVILKVILSKLVKPEGIPITPMASAIGRGLEDEIRFGRIRDKEK 172
YP_009044255.1            KAVKDRSLRISCESAAVILKVILSKLVKPEGIPITPMASAIGRGLEDEIRFGRIRDKEK 172
AEH41021.1                 AEVLNEKRGVSSEALITLLIKVVLGTLTLDASKATIQQVSSQLGKALEDEARFGRIRREQA 179
sp|P18147|RPOL BPK11      AEVLNEKRGVSSEALITLLIKVVLGTLTLDASKATIQQVSSQLGKALEDEARFGRIRREQA 179
ACY75835.1                 -----IKPEAVAYITIKTTLACLTSADNTTVQAVASAIGRAIEDEARFGRIRDLEA 159
AC057213.1                 -----IKPEAVAYITIKTTLACLTSADNTTVQAVASAIGRAIEDEARFGRIRDLEA 159
tr|C6ZCU5|C6ZCU5_LAMB    -----IKPEAVAYITIKTTLACLTSADNTTVQAVASAIGRAIEDEARFGRIRDLEA 159
CAC86264.1                 -----LKPEASAFITLKVILASLTSTNMTTIQAAAGMLGKAIEDEARFGRIRDLEA 160
: . * * : : * . : : : : : : : : : : : * : * : * : * :
sp|P06221|RPOL_BPSP6      KYFEKVKKS-LKASRTKSYRHAHNVAVVAEKSVAEKDADFDRWEAWPKETQLQIGITLLE 188
AAZ72968.1                 EHFKKAIADNLNKRAGASYKKA Y-MQAVEASMLEQGQLE-DAWGTWSPTEAVHVGIKMLE 230
YP_009044255.1            EHFKKAIADNLNKRAGASYKKA Y-MQAVETSMLEQGQLE-DAWGTWSPTEAVHVGIKMLE 230
AEH41021.1                 SYFKKNVADQLDKRVGHVYKKA F-MQVVEADMI SKGMLGGDNWSWKTE QMHVGTKLE 238
sp|P18147|RPOL BPK11      AYFKKNVADQLDKRVGHVYKKA F-MQVVEADMI SKGMLGGDNWSWKTE QMHVGTKLE 238
ACY75835.1                 KHFKKNVEEQLNKRQGHVYKKA F-MQVVEADMLS LKGGLGGEAWSSWKHEDSIHVGVRCEI 218
AC057213.1                 KHFKKNVEEQLNKRQGHVYKKA F-MQVVEADMLS LKGGLGGEAWSSWKHEDSIHVGVRCEI 218
tr|C6ZCU5|C6ZCU5_LAMB    KHFKKNVEEQLNKRQGHVYKKA F-MQVVEADMLS LKGGLGGEAWSSWKHEDSIHVGVRCEI 218
CAC86264.1                 KHFKKNVEEQLNKRQGHVYKKA F-MQVVEADMIGRLLGGEAWSSWDKETIMHVGI RLEI 219
: * * : . * . : * * : : * . : : : : : : : : : * : * : * : * :
sp|P06221|RPOL_BPSP6      ILEGSVFYNGEPVFMFRMARTYGGKTIYYLQTSSESVGQW---SAFKEHVAQLSPAYAPCV 245
AAZ72968.1                 IVIQTSTQVLE---LKRYGAGNAAD---VEMVHLSDFWVKKMAQRGFSLAGIAPVYDPCV 284
YP_009044255.1            IVIQTSTQVLE---LKRYGAGNAATD---VEMVHLSDFWVKKMAQRGFSLAGIAPVYDPCV 284
AEH41021.1                 LLIEGTGLVE---MTKNKMADGSDDDVTSMQMVLAPAFVELLSKRALAGISPMYDPCV 295
sp|P18147|RPOL BPK11      LLIEGTGLVE---MTKNKMADGSDDDVTSMQMVLAPAFVELLSKRALAGISPMYDPCV 295
ACY75835.1                 MLIESTGMVS---LHRQAGVVQGD---SEI IELAPEYAEAIATRAGALAGISPMFOPCV 272
AC057213.1                 MLIESTGMVS---LHRQAGVVQGD---SEI IELAPEYAEAIATRAGALAGISPMFOPCV 272
tr|C6ZCU5|C6ZCU5_LAMB    MLIESTGMVS---LHRQAGVVQGD---SEI IELAPEYAEAIATRAGALAGISPMFOPCV 272
CAC86264.1                 MLIESTGLVE---LQRHNAGNAGSD---HEALQLAQEYVDVLAKRAGALAGISPMFOPCV 273
: : . . : : : : : : : : : : . * : : : : * * :
sp|P06221|RPOL_BPSP6      IPRRPRTPFNGGFFTEKVASRIRLVKGNREHVRKLTQRMPKVKYKAINALQNTQWQINK 305
AAZ72968.1                 VPKPKWTGVVGGGYAKGRRPLPLIRLGSKSAVARYEDVYMPVEVYEAIVNIQNTPWKVNK 344
YP_009044255.1            VPKPKWTGVVGGGYAKGRRPLPLIRLGSKSAVARYEDVYMPVEVYEAIVNIQNTPWKVNK 344
AEH41021.1                 VPKPKWYETVGGGYVSGRRPLALVRTHSKKALRRYEDVHMPEVYKAVNLAQNTPWKVNK 355
sp|P18147|RPOL BPK11      VPKPKWYETVGGGYVSGRRPLALVRTHSKKALRRYEDVHMPEVYKAVNLAQNTPWKVNK 355
ACY75835.1                 VPKPKWTGITGGGYWANGRRPLALVRTHSKKALMRYEDVYMPVEVYKAINIQNTAWKINK 332
AC057213.1                 VPKPKWTGITGGGYWANGRRPLALVRTHSKKALMRYEDVYMPVEVYKAINIQNTAWKINK 332
tr|C6ZCU5|C6ZCU5_LAMB    VPKPKWTGITGGGYWANGRRPLALVRTHSKKALMRYEDVYMPVEVYKAINIQNTAWKINK 332
CAC86264.1                 VPKPKWYAITGGGYWANGRRPLALVRTHSKKGLMRYEDVYMPVEVYKAVNLAQNTAWKINK 333
: * * * : * * : : : : : : : : * * * : * * :
sp|P06221|RPOL_BPSP6      DVLAVIEEVIRLDLGYGVPFSLDKENKPANVPVVEFQHLRGLRELKEMLSPEQWQFPI 365
AAZ72968.1                 KVLDDVNMVVKLNNTIP--IDDI PQMEPL----KP--EAYA-----GETEELK 384
YP_009044255.1            KVLDDVNMVVKLNNTIP--IDDI PQMEPL----KP--EDYA-----GETEELK 384
AEH41021.1                 KVLAVVNEIINWKHCP--VGDVPAIEREELPPRP--DD-I-----DTNEVARK 398
sp|P18147|RPOL BPK11      KVLAVVNEIINWKHCP--VGDVPAIEREELPPRP--DD-I-----DTNEVARK 398
ACY75835.1                 KVLAVANVITKWKHCPC--VEDI PAIEREELPMKP--ED-I-----DMNPEALT 375
AC057213.1                 KVLAVANVITKWKHCPC--VEDI PAIEREELPMKP--ED-I-----DMNPEALT 375
tr|C6ZCU5|C6ZCU5_LAMB    KVLAVANVITKWKHCPC--VEDI PAIEREELPMKP--ED-I-----DMNPEALT 375
CAC86264.1                 KVLAVVNEIIVNWKNCPC--VADI PSLERQELPPKP--DD-I-----DTNEAALK 376
* * * : : . * : : : *
sp|P06221|RPOL_BPSP6      NWKGECARLYTAETKRGSKSAAVVRMVGQARKYSAFESIYFVYAMDSSRVTYVQSSTLSP 425
AAZ72968.1                 AWKKAAGIYRREKARQRSRLLSFI VQANKFSQFKAIWFFPYNMDWGRVYAV-PMFNP 443
YP_009044255.1            AWKKAAGIYRREKARQRSRLLSFI VQANKFSQFKAIWFFPYNMDWGRVYAV-PMFNP 443
AEH41021.1                 AWKRAAAA VYRKDKARQRSRLSMFMVAQANKFANHKA IWFFPYNMDWGRVYAV-SMFNP 457
sp|P18147|RPOL BPK11      AWKRAAAA VYRKDKARQRSRCRCEFMVAQANKFANHKA IWFFPYNMDWGRVYAV-SMFNP 457
ACY75835.1                 AWKRAAAA VYRKDKARQRSRISLEFMLEQANKFANHKA IWFFPYNMDWGRVYAV-SMFNP 434
AC057213.1                 AWKRAAAA VYRKDKARQRSRISLEFMLEQANKFANHKA IWFFPYNMDWGRVYAV-SMFNP 434
tr|C6ZCU5|C6ZCU5_LAMB    AWKRAAAA VYRKDKARQRSRISLEFMLEQANKFANHKA IWFFPYNMDWGRVYAV-SMFNP 434
CAC86264.1                 EWKKAAGIYRLDKARVSRISLEFMLEQANKFASKKAIWFFPYNMDWGRVYAV-PMFNP 435
* : * * : * * : : : : : : : : * * * : * * : *
```

```

sp|P06221|RPOL_BPSP6      QSNDLGKALLRFTEGRFVNGVEALKWFG  

AAZ72968.1                IINGANLWGWDDKKTFDVVRVSNVLD  

YP_009044255.1          EEFQDMCRD 485
AEH41021.1              QGNDMQKGLLTLAVGKPI--GADGFKWLVH  

                                GANCAGVDKVTFEERIKWVE  

sp|P18147|RPOL BPK11     QGNDMQKGLLTLAVGKPI--GADGFKWLVH  

AC Y75835.1              GANCAGVDKVTFEERIKWVE  

AC057213.1              DNHD--NIMA 500
tr|C62CU5|C62CU5_LAMBD  QGNDMTKGLLTLAKGKPI--GLDGYFWLKI  

CAC86264.1              IHGANCAGVDKVPFPERIKFIE  

                                ENEA--NILA 514
                                QGNDMTKGLLTLAKGKPI--GLDGYFWLKI  

                                IHGANCAGVDKVPFPERIKFIE  

                                ENEG--NILA 514
                                QGNDMTKGLLTLAKGKPI--GKEGYFWLKI  

                                IHGANCAGVDKVPFPERIKFIE  

                                ENHE--NIMA 491
                                QGNDMTKGLLTLAKGKPI--GKEGYFWLKI  

                                IHGANCAGVDKVPFPERIKFIE  

                                ENHE--NIMA 491
                                QGNDMTKGLLTLAKGKPI--GKEGYFWLKI  

                                IHGANCAGVDKVPFPERIKFIE  

                                ENHE--NIMA 491
                                QGNDMTKGLLTLAKGKPI--GEGGYFWLKI  

                                IHGANCAGVDKVPFPERIAFIE  

                                KHVD--DILA 492
                                * . * * : : * : * : * : * : * : * : * : * : * : * : * :

```

```

sp|P06221|RPOL_BPSP6      IAADPLTF-TQWAKADAPYEF  

AAZ72968.1                FLAWCFEYAYLDLVDEGRADEFRTHL  

YP_009044255.1          FVHFDGSCSGIQH 544
AEH41021.1              AAKAPMDSIEWWGLDSPFCFLAFCFEYAV  

                                VMH----HGLSYSCSLPIA  

sp|P18147|RPOL BPK11     AAKAPMDSIEWWGLDSPFCFLAFCFEYAV  

AC Y75835.1              VMH----HGLSYSCSLPIA  

AC057213.1              FDGSCSGIQH 555
tr|C62CU5|C62CU5_LAMBD  SAADPLNN-TWWTQDSSPFCFLAFCFEYAV  

CAC86264.1              FVHK----HGLNYNCSLPLA  

                                FDGSCSGIQH 568
                                SAADPLNN-TWWTQDSSPFCFLAFCFEYAV  

                                FVHK----HGLNYNCSLPLA  

                                FDGSCSGIQH 568
                                CAKSPLEN-TWVAEQDSSPFCFLAFCFEYAV  

                                FVQH----HGLSYNCSLPLA  

                                FDGSCSGIQH 545
                                CAKSPLEN-TWVAEQDSSPFCFLAFCFEYAV  

                                FVQH----HGLSYNCSLPLA  

                                FDGSCSGIQH 545
                                CAKSPLEN-TWVAEQDSSPFCFLAFCFEYAV  

                                FVQH----HGLSYNCSLPLA  

                                FDGSCSGIQH 545
                                CAKDPINN-TWVAEQDSSPFCFLAFCFEYAV  

                                FVTH----HGLSYNCSLPLA  

                                FDGSCSGIQH 546
                                * * : * : * : * : * : * : * : * : * : * : * : * : * : * :

```

```

sp|P06221|RPOL_BPSP6      YSAMPLRDHVGA  

AAZ72968.1                KAVNLLKPSDA  

YP_009044255.1          QDIYGA  

AEH41021.1              VAQVVIKKNALY-----MDADDATTFT 594
sp|P18147|RPOL BPK11     FSAMPLRDHIGGHAVNLL  

AC Y75835.1              P  

AC057213.1              SGKVDIYRIVSDRI  

tr|C62CU5|C62CU5_LAMBD  FSAMPLRDHIGGHAVNLL  

CAC86264.1              P  

                                SGKVDIYRIVSDRI  

                                EEE  

                                LKVVLLVNGTDNEMV  

                                THEDKKTGEI 615
                                FSAMPLRDHIGGHAVNLL  

                                P  

                                SGKVDIYRIVSDRI  

                                EEE  

                                LKVVLLVNGTDNEMV  

                                THEDKKTGEI 615
                                FSAMPLRDIIGGRAVNLL  

                                P  

                                SDTVQDIYKIVADK  

                                VNEVLHQQHVIN  

                                GSIQVVEQIADK  

                                ETGEF 628
                                FSAMPLRDIIGGRAVNLL  

                                P  

                                SDTVQDIYKIVADK  

                                VNEVLHQQHVIN  

                                GSIQVVEQIADK  

                                ETGEF 628
                                FSAMPLRDHVGGRAVNLL  

                                P  

                                SETVQDIYGI  

                                VAKKVI  

                                ELLQDAINGTD  

                                NEVTVTDENTGEI 605
                                FSAMPLRDHVGGRAVNLL  

                                P  

                                SETVQDIYGI  

                                VAKKVI  

                                ELLQDAINGTD  

                                NEVTVTDENTGEI 605
                                FSAMPLRDHVGGRAVNLL  

                                P  

                                SETVQDIYGI  

                                VAKKVI  

                                ELLQDAINGTD  

                                NEVTVTDENTGEI 605
                                FSAMPLRDHVGGRAVNLL  

                                P  

                                SETVQDIYGI  

                                VAKKVI  

                                ELLQDAINGTD  

                                NEVTVTDENTGEI 605
                                FSAMPLRDHVGGRAVNLL  

                                P  

                                SETVQDIYGI  

                                VAKKVI  

                                ELLQDAINGTD  

                                PNEMI  

                                TVTDKDTGEI 606
                                ***** : * : * * : * : * : * : * : * : * : * : * : * :

```

```

sp|P06221|RPOL_BPSP6      SGSVTLSGTEL  

AAZ72968.1                RAMASAWDS  

YP_009044255.1          TGIITRSL  

AEH41021.1              TKK  

sp|P18147|RPOL BPK11     TERLK---LGTRELARQWL  

AC Y75835.1              T  

AC057213.1              YGSRKVT  

tr|C62CU5|C62CU5_LAMBD  TERLK---LGTRELARQWL  

CAC86264.1              T  

                                YGSRKVT  

                                TRRSVMTL  

                                YG$KEYGFADQVVE  

                                DIVMP----- 667
                                TERLK---LGTRELARQWL  

                                T  

                                YGSRKVT  

                                TRRSVMTL  

                                YG$KEYGFADQVVE  

                                DIVMP----- 667
                                REKVM---LGSVLA  

                                AQLQYGV  

                                TRKVT  

                                TRRSVMTL  

                                YG$KEFGFRQVLE  

                                DTIQP----- 680
                                HEKVT---LGSVLA  

                                AQLQYGV  

                                TRKVT  

                                TRRSVMTL  

                                YG$KESLVRQVLE  

                                DTIQP----- 680
                                SEKVK---LGT  

                                KALAGQWLA  

                                YGVT  

                                TRSV  

                                TRRSVMTL  

                                YG$KEFGFRQVLE  

                                DTIQP----- 657
                                SEKVK---LGT  

                                KALAGQWLA  

                                YGVT  

                                TRSV  

                                TRRSVMTL  

                                YG$KEFGFRQVLE  

                                DTIQP----- 657
                                SEKVK---LGT  

                                KALAGQWLA  

                                YGVT  

                                TRSV  

                                TRRSVMTL  

                                YG$KEFGFRQVLE  

                                DTIQP----- 657
                                SEKLK---LGT  

                                STLAQWLA  

                                YGVT  

                                TRSV  

                                TRRSVMTL  

                                YG$KEFGFRQVLD  

                                DTIQP----- 658
                                : * : * : * : * : * : * : * : * : * : * : * :

```

```

sp|P06221|RPOL_BPSP6      QKAVAEGRTANKVHP  

AAZ72968.1                FEDDRQDYLT  

YP_009044255.1          PGAA  

AEH41021.1              NYMTALIWPSI  

sp|P18147|RPOL BPK11     --AIDSGS-----GAM  

AC Y75835.1                FTEPSQASRFMA  

AC057213.1                KMIWEAVSVTV  

tr|C62CU5|C62CU5_LAMBD  --AIDSGS-----GAM  

CAC86264.1                FTEPSQASRFMA  

                                KMIWEAVSVTV  

                                VAAV  

                                DAMKWLQGA  

                                AAK 714
                                --AIDSGS-----GLM  

                                FTHPNQAA  

                                GYMAKL  

                                IWD  

                                AVTV  

                                VVA  

                                AVE  

                                AMNWLKSA  

                                AAK 727
                                --AIDNGE-----GLM  

                                FTHPNQAA  

                                GYMAKL  

                                IWD  

                                AVTV  

                                VVA  

                                AVE  

                                AMNWLKSA  

                                AAK 727
                                --AIDSGK-----GLM  

                                FTQPNQAA  

                                GYMAKL  

                                IWES  

                                VSVTV  

                                VAAVE  

                                AMNWLKSA  

                                AAK 704
                                --AIDSGK-----GLM  

                                FTQPNQAA  

                                GYMAKL  

                                IWES  

                                VSVTV  

                                VAAVE  

                                AMNWLKSA  

                                AAK 704
                                --AIDSGK-----GLM  

                                FTQPNQAA  

                                GYMAKL  

                                IWES  

                                VSVTV  

                                VAAVE  

                                AMNWLKSA  

                                AAK 704
                                --AIDSGK-----GLM  

                                FTQPNQAA  

                                GYMAKL  

                                IWD  

                                AVSV  

                                TVV  

                                AA  

                                VE  

                                AMNWLKSA  

                                AAK 705
                                * : * * : * : * : * : * : * : * : * : * : * :

```

```

sp|P06221|RPOL_BPSP6      FAAK-----RNEGLM  

AAZ72968.1                YTLPTGF  

YP_009044255.1          LEQKIMATE  

AEH41021.1              MLRVRT  

sp|P18147|RPOL BPK11     LLAAEVKDKKTGEI  

AC Y75835.1                LKPCLPVH  

AC057213.1                VWT  

tr|C62CU5|C62CU5_LAMBD  LLAAEVKDKKTGEI  

CAC86264.1                LKPCLPVH  

                                VWT  

                                PDGFFPV  

                                WQY  

                                YR  

                                KKDT  

                                TR  

                                LNMFL  

                                GSNL  

                                QPT  

                                VNK  

                                G 774
                                LLAAEVKDKKTGEI  

                                LKPCLPVH  

                                VWT  

                                PDGFFPV  

                                WQY  

                                YR  

                                KKDT  

                                TR  

                                LNMFL  

                                GSNL  

                                QPT  

                                VNK  

                                G 774
                                LLAAEVKDKKTGEI  

                                LKPCLPVH  

                                VWT  

                                PDGFFPV  

                                WQY  

                                YR  

                                KKDT  

                                TR  

                                LNMFL  

                                GSNL  

                                QPT  

                                VNK  

                                G 787
                                LLAAEVKDKKTGEI  

                                LKPCLPVH  

                                VWT  

                                PDGFFPV  

                                WQY  

                                YR  

                                KKDT  

                                TR  

                                LNMFL  

                                GSNL  

                                QPT  

                                VNK  

                                G 787
                                LLAAEVKDKKTGEI  

                                LKPCLPVH  

                                VWT  

                                PDGFFPV  

                                WQY  

                                YR  

                                KKDT  

                                TR  

                                LNMFL  

                                GSNL  

                                QPT  

                                VNK  

                                G 764
                                LLAAEVKDKKTGEI  

                                LKPCLPVH  

                                VWT  

                                PDGFFPV  

                                WQY  

                                YR  

                                KKDT  

                                TR  

                                LNMFL  

                                GSNL  

                                QPT  

                                VNK  

                                G 764
                                LLAAEVKDKKTGEI  

                                LKPCLPVH  

                                VWT  

                                PDGFFPV  

                                WQY  

                                YR  

                                KKDT  

                                TR  

                                LNMFL  

                                GSNL  

                                QPT  

                                VNK  

                                G 764
                                LLAAEVKDKKTGEI  

                                LKPCLPVH  

                                VWT  

                                PDGFFPV  

                                WQY  

                                YR  

                                KKDT  

                                TR  

                                LNMFL  

                                GSNL  

                                QPT  

                                VNK  

                                G 765
                                : * : * : * : * : * : * : * : * : * : * : * :

```

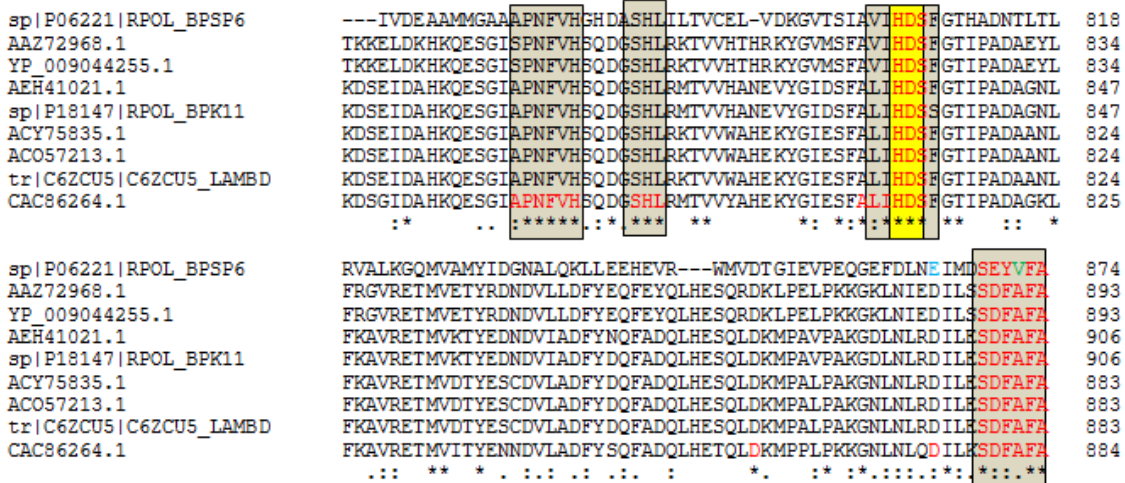


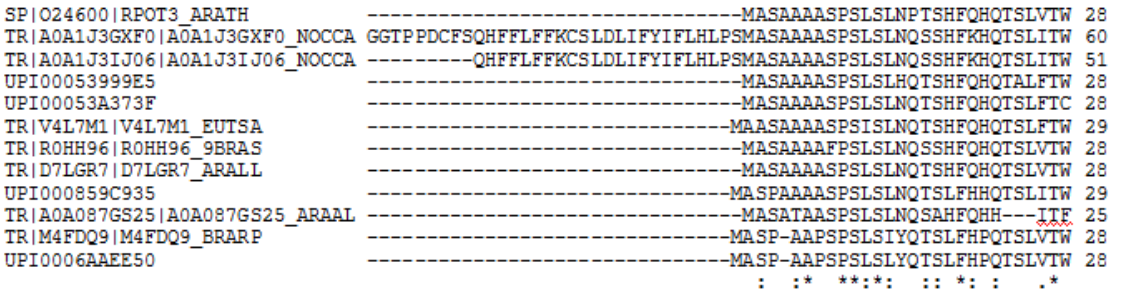
Fig. 2. MSA of SSU RNA polymerases from enterobacteria phages like T3, T7, SP6, K11, λ, etc

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 LKPPSS--ALFRKTLPPFERHSLPIS--ASS-SS--SSSSTSLVHEKPISNSVHFH 81
 TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA LKPPSS--ALFRKTLPLSSPVR---RLSLPI SASSSS--SSSSTSLVSEKPTANSVHFH 114
 TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA LKPPSS--ALFRKTLPLSSPVR---RLSLPI SASSSSSSSSTSLVSEKPTANSVHFH 106
 UPI00053999E5 LKPPSSY--ALFRKTLPPFERHSLPIS--ASS-SSSSSSSSTSLVHEKPISNSVHFH 83
 UPI00053A373F LKPPSSP--ALFRKTLPPFERHSLPIS--ASS-SSSSSSSSTSLVHEKPISNSVHFH 83
 TR|V4L7M1|V4L7M1_EUTSA LKPPSS---ALFRKTLPPFDPSSPKRISLPP-ISASSSSSSASLSVSEKPT---TVHFH 82
 TR|ROHH96|ROHH96_9BRAS LKPP-SS--ALYRRKTLPPFERHSLPIS--ASS-SS---SSSSTSLVHEKPISNSVHFH 79
 TR|D7LGR7|D7LGR7_ARALL LKPPSS--ALFRKTLPPFERHSLPIS--ASS-SS-----SSSSTSLVHEKPISNSVHFH 78
 UPI000859C935 LKPPSSSSALFRKATKRL-----PPI-SAASSS--SSTLSV--TTEKPTVHFH 76
 TR|A0A087GS25|A0A087GS25_ARAAL LKPTSS--TLRRLKTLPPSVKR-----ISASSSSSSSSTSLVTEKPTNSVHFH 75
 TR|M4FDQ9|M4FDQ9_BRARP LKPPSS--SALFRKTKRLL-----PPI-SAASSSSSSTLSV--TEKPTVHFH 74
 UPI0006AAEE50 LKPPSS--SALFRKTKRLL-----PPI-SAASSSSSSTLSV--TEKPTVHFH 74
 **** * : * *** : ** : * : ****

SP|O24600|RPOT3_ARATH GNLESFENQDSSYAGTIKASLIEELENPVERNGLSGRRRLFMQDP PWISALFLKGLSK 141
 TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA GNLESFENQD-SFAGSINGTSLIDELENPQVNGISGRRRLFMQDP PWISALFLKGLSK 173
 TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA GNLESFENQD-SFAGSINGTSLIDELENPQVNGISGRRRLFMQDP PWISALFLKGLSK 165
 UPI00053999E5 GNLESFENQDSSYAGTIKASLIEELENPVERNGLSGRRRLFMQDP PWISALFLKGLSK 143
 UPI00053A373F GNLESFENQDSSYAGTIKASLIEELENPVERNGLSGRRRLFMQDP PWISALFLKGLSK 143
 TR|V4L7M1|V4L7M1_EUTSA GNLESFENQD-SFAGTNGTSLIEELENPVERNGLSGRRRLFMQDP PWISALFLKGLSK 141
 TR|ROHH96|ROHH96_9BRAS GNLESFENQDSSYAGTIKASLIEELENPVERNGLSGRRRLFMQDP PWISALFLKGLSK 139
 TR|D7LGR7|D7LGR7_ARALL GNLESFENQDSSYAGTIKASLIEELENPVERNGLSGRRRLFMQDP PWISALFLKGLSK 138
 UPI000859C935 GNLESFESH----AGTIKGFAT---TNPVERNELSARKRLFTQDP PWISALFLKGLTK 129
 TR|A0A087GS25|A0A087GS25_ARAAL GNLEDSFENQSG----TIKGATL---IENPVER SELSGRRRLFMQDP PWISALFLKGLTK 128
 TR|M4FDQ9|M4FDQ9_BRARP GNLESFESH-D-SFPGAIKGAAF---TENPVER SELSATRRLFTQDP PWISALFLKGLTK 130
 UPI0006AAEE50 GNLESFEGHD-SFPGAIKGAAF---TENPVER SELSATRRLFTQDP PWISALFLKGLTK 130
 **** : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

SP|O24600|RPOT3_ARATH MV-DQTLKIERKIDIKRKFDSLRRRQVKEETEAWERMVDEYRDLKEKCEKNLAPNLPYV 200
 TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA MV-DQTLKIERKIDIKRKFDSLRRRQVKEETEAWERMVDEYRDLKEKCEKNLAPNLPYV 232
 TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA MV-DQTLKIERKIDIKRKFDSLRRRQVKEETEAWERMVDEYRDLKEKCEKNLAPNLPYV 224
 UPI00053999E5 MV-DQTVKIEHKIDIKRKFDSLRRRQVKEETEAWERMVDEYRDLKEKCEKNLAPNLPYV 202
 UPI00053A373F MV-DQTVKIEHKIDIKRKFDSLRRRQVKEETEAWERMVDEYRDLKEKCEKNLAPNLPYV 202
 TR|V4L7M1|V4L7M1_EUTSA MV-DQTVKIERKIDIKRKFDSLRRRQVKEETEAWERMVDEYRDLKEKCEKNLAPNLPYV 200
 TR|ROHH96|ROHH96_9BRAS MV-DQTVKIEHKIDIKRKFDSLRRRQVKEETEAWERMVDEYRDLKEKCEKNLAPNLPYV 198
 TR|D7LGR7|D7LGR7_ARALL MV-DQTLKIERKIDIKRKFDSLRRRQVKEETEAWERMVDEYRDLKEKCEKNLAPNLPYV 197
 UPI000859C935 MAADQTVKIERKIDIKRKFDSLRRRQVKEETEAWERTVDEYRDLKEKCEKSLAPSLPV 189
 TR|A0A087GS25|A0A087GS25_ARAAL MV-DQTFQIERKIDIKRKFDSLRRRQVKEETEAWERMVDEYRDLKEKCEKNLAPNLPYV 187
 TR|M4FDQ9|M4FDQ9_BRARP M---TVKIERKIDIKRKFDSLRRRQVKEETEAWERMVDEYRDLKEKCEKSLAPNLPYV 186
 UPI0006AAEE50 M---TVKIERKIDIKRKFDSLRRRQVKEETEAWERMVDEYRDLKEKCEKSLAPNLPYV 186
 * * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

SP|O24600|RPOT3_ARATH KHMFLGWFPQPLKVDIEREQKIQKNKSKKVRVAAAPHIELLPADKMAVIVMHKMMGLVMSG 260
 TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA KHMFLGWFPQPLKVDIEREQKIQKNKSKKVRVAAAPHIELLPADKMAVIVMHKMMGLVMSG 292
 TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA KHMFLGWFPQPLKVDIEREQKIQKNKSKKVRVAAAPHIELLPADKMAVIVMHKMMGLVMSG 284
 UPI00053999E5 KHMFLGWFPQPLKVDIEREQKIQKNKSKKVRVAAAPHIELLPADKMAVIVMHKMMGLVMSG 262
 UPI00053A373F KHMFLGWFPQPLKVDIEREQKIQKNKSKKVRVAAAPHIELLPADKMAVIVMHKMMGLVMSG 262
 TR|V4L7M1|V4L7M1_EUTSA KHMFLGWFPQPLKVDIEREQKIQKNKSKKVRVAAAPHIELLPADKMAVIVMHKMMGLVMSG 260
 TR|ROHH96|ROHH96_9BRAS KHMFLGWFPQPLKVDIEREQKIQKNKSKKVRVAAAPHIELLPADKMAVIVMHKMMGLVMSG 258
 TR|D7LGR7|D7LGR7_ARALL KHMFLGWFPQPLKVDIEREQKIQKNKSKKVRVAAAPHIELLPADKMAVIVMHKMMGLVMSG 257
 UPI000859C935 KHMFLGWFPQPLKVDIEREQKIQKNKSKKVRVAAAPHIELLPADKMAVIVMHKMMGLVMSG 249
 TR|A0A087GS25|A0A087GS25_ARAAL KHMFLGWFPQPLKVDIEREQKIQKNKSKKVRVAAAPHIELLPADKMAVIVMHKMMGLVMSG 247
 TR|M4FDQ9|M4FDQ9_BRARP KHMFLGWFPQPLKVDIEREQKIQKNKSKKVRVAAAPHIELLPADKMAVIVMHKMMGLVMSG 246
 UPI0006AAEE50 KHMFLGWFPQPLKVDIEREQKIQKNKSKKVRVAAAPHIELLPADKMAVIVMHKMMGLVMSG 246
 ***** : ***** : ***** : ***** : ***** : *****

SP|O24600|RPOT3_ARATH HEDGCIQVVQAAVSGIATIEQEVRIHNFLKTRKNNAGDSQEELKEKQLLRKRVNSLIRR 320
 TR|A0A1J3GXF0|A0A1J3GXF0_NOCCA HEDGCIQVVQAAVSGIATIEQEVRIHNFLKTRKNNAGDSQEELKDKQLLRKRVNSLIRR 352
 TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA HEDGCIQVVQAAVSGIATIEQEVRIHNFLKTRKNNAGDSQEELKDKQLLRKRVNSLIRR 344
 UPI00053999E5 HEDGCIQVVQAAVSGIATIEQEVRIHNFLKTRKNNAGDSQEELKEKQLLRKRVNSLIRR 322
 UPI00053A373F HEDGCIQVVQAAVSGIATIEQEVRIHNFLKTRKNNAGDSQEELKEKQLLRKRVNSLIRR 322
 TR|V4L7M1|V4L7M1_EUTSA HEDGCIQVVQAAVSGIATIEQEVRIHNFLKTRKNNAGDSEELKDKQLLRKRVNSLIRR 320
 TR|ROHH96|ROHH96_9BRAS HEDGCIQVVQAAVSGIATIEQEVRIHNFLKTRKNNAGDSQEELKEKQLLRKRVNSLIRR 318
 TR|D7LGR7|D7LGR7_ARALL HEDGCIQVVQAAVSGIATIEQEVRIHNFLKTRKNNAGDSQEELKEKQLLRKRVNSLIRR 317
 UPI000859C935 HEDGCIQVVQAAVSGIATIEQEVRIHNFLKTRKNNAGDSEELKDKQLLRKRVNSLIRR 309
 TR|A0A087GS25|A0A087GS25_ARAAL HEDGCIQVVQAAVSGIATIEQEVRIHNFLKTRKNNAGDTEELKDKQLLRKRVNSLIRR 307
 TR|M4FDQ9|M4FDQ9_BRARP HEDGCIQVVQAAVSGIATIEQEVRIHNFLKTRKNNAGDSQEELKDKQLLRKRVNSLIRR 306
 UPI0006AAEE50 HEDGCIQVVQAAVSGIATIEQEVRIHNFLKTRKNNAGDSQEELKDKQLLRKRVNSLIRR 306
 ***** : ***** : ***** : ***** : ***** : *****

SP|O24600|RPOT3_ARATH KRII DALKVVKSEGTKEPWRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 380
 TR|A0A1J3GXFO|A0A1J3GXFO_NOCCA KRMI DALKVVKCEGKIPWGRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 412
 TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA KRMI DALKVVKCEGKIPWGRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 404
 UPI00053999E5 KRII DALKVVKSEGTKEPWRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 382
 UPI00053A373F KRII DALKVVKSEGTKEPWRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 382
 TR|V4L7M1|V4L7M1_EUTSA KRII DALKVVKSEGTKEPWRATQAKLGSRLLELLIETAYVQPPLTQAGDTIPEFRPAFRH 380
 TR|R0HH96|R0HH96_9BRAS KRII DALKVVKSEGTKEPWRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 378
 TR|D7LGR7|D7LGR7_ARALL KRII DALKVVKSEGTKEPWRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 377
 UPI000859C935 KRII DALKVVKSEGTKEPWRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 369
 TR|A0A087GS25|A0A087GS25_ARAAL KRII DALKVVKCEGKIPWGRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 367
 TR|M4FDQ9|M4FDQ9_BRARP KRII DALKVVKSEGTKEPWRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 366
 UPI0006AAEE50 KRII DALKVVKSEGTKEPWRATQAKLGSRLLELLIETAYVQPPLTQSGDSIPEFRPAFRH 366
 :*** * *****:*****:*****:*****:*****

SP|O24600|RPOT3_ARATH RFKTVTKYPGSKLVRVYVIECDLSLLLAGLDKSAKHMLIPYVPMVLP PKRWKGYDKGGYL 440
 TR|A0A1J3GXFO|A0A1J3GXFO_NOCCA KFKTVTKYPGSKLVRVYVIECDLSLLLAGLDKSAKHMLIPYVPMVLP PKRWKGYDKGGYL 472
 TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA KFKTVTKYPGSKLVRVYVIECDLSLLLAGLDKSAKHMLIPYVPMVLP PKRWKGYDKGGYL 464
 UPI00053999E5 TFKTVTKYPGSKLVRVYVIECDLSLLLAGLDKSAKHMLIPYVPMVLP PKRWKGYDKGGYL 442
 UPI00053A373F UPI00053A373F TFKTVTKYPGSKLVRVYVIECDLSLLLAGLDKSAKHMLIPYVPMVLP PKRWKGYDKGGYL 442
 TR|V4L7M1|V4L7M1_EUTSA KFKTVTKYPGSKLVRVYVIECDLSLLLAGLDKSAKHMLIPYVPMVLP PKRWKGYDKGGYL 440
 TR|R0HH96|R0HH96_9BRAS KFKTVTKYPGSKLVRVYVIECDLSLLLAGLDKSAKHMLIPYVPMVLP PKRWKGYDKGGYL 438
 TR|D7LGR7|D7LGR7_ARALL KFKTVTKYPGSKLVRVYVIECDLSLLLAGLDKSAKHMLIPYVPMVLP PKRWKGYDKGGYL 437
 UPI000859C935 KFKTVTKYPGSKLVRVYVIECDLSLLLAGLDKSAKHMLIPYVPMVLP PKRWKGYDKGGYL 429
 TR|A0A087GS25|A0A087GS25_ARAAL KFKTVTKYPGSKLVRVYVIECDLSLLLAGLDKSAKHMLIPYVPMVLP PKRWKGYDKGGYL 427
 TR|M4FDQ9|M4FDQ9_BRARP KFKTVTKYPGSKLVRVYVIECDLSLLLAGLDKSAKHMLIPYVPMVLP PKRWKGYDKGGYL 426
 UPI0006AAEE50 KFKTVTKYPGSKLVRVYVIECDLSLLLAGLDKSAKHMLIPYVPMVLP PKRWKGYDKGGYL 426
 *****:*****:*****:*****:*****

SP|O24600|RPOT3_ARATH FLPSYIMRTHGSKKQDQALDIDSHKTAHRVFEALDTLGNTKWRVNRNILDVVERLWADGG 500
 TR|A0A1J3GXFO|A0A1J3GXFO_NOCCA FLPSYIMRTHGSKKQDQALDIDSSKTAHRVFEALDTLGNTKWRVNRKILDVVEKLWADGG 532
 TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA FLPSYIMRTHGSKKQDQALDIDSSKTAHRVFEALDTLGNTKWRVNRKILDVVEKLWADGG 524
 UPI00053999E5 FLPSYIMRTHGSKKQDQALDIDSYKTAHRVFEALDTLGNTKWRVNRKILDVVERLWADGG 502
 UPI00053A373F FLPSYIMRTHGSKKQDQALDIDSHKTAHRVFEALDTLGNTKWRVNRKILDVVERLWADGG 502
 TR|V4L7M1|V4L7M1_EUTSA FLPSYIMRTHGSKKQDQALDIDSSKTAHRVFEALDTLGNTKWRVNRKILDVVERLWADGG 500
 TR|R0HH96|R0HH96_9BRAS FLPSYIMRTHGSKKQDQALDIDSSKTAHRVFEALDTLGNTKWRVNRKILDVVERLWADGG 498
 TR|D7LGR7|D7LGR7_ARALL FLPSYIMRTHGSKKQDQALDIDSHKTAHRVFEALDTLGNTKWRVNRNILDVVERLWADGG 497
 UPI000859C935 FLPSYIMRTHGSKKQDQALDIDSSKTAHRVFEALDTLGNTKWRVNRKILDVVERLWADGG 489
 TR|A0A087GS25|A0A087GS25_ARAAL FLPSYIMRTHGSKKQDQALDIDVSHKTAHRVFEALDTLGNTKWRVNRNILDVVERLWADGG 487
 TR|M4FDQ9|M4FDQ9_BRARP FLPSYIMRTHGSKKQDQALDIDSSKTAHRVFEALDTLGNTKWRVNRKILDVVERLWADGG 486
 UPI0006AAEE50 FLPSYIMRTHGSKKQDQALDIDSSKTAHRVFEALDTLGNTKWRVNRKILDVVERLWADGG 486
 *****:*****:*****:*****:*****

SP|O24600|RPOT3_ARATH NIAGLVNREDVPIPEKPSSEDPPELQSWKWSARKANKINRERHSLRC DVELKLSVARMKM 560
 TR|A0A1J3GXFO|A0A1J3GXFO_NOCCA NIAGLVNREDVPIPEKPTS EDPEEMQSWKWSARKANKINRERHSLRC DVELKLSVARMKM 592
 TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA NIAGLVNREDVPIPEKPTS EDPEEMQSWKWSARKANKINRERHSLRC DVELKLSVARMKM 584
 UPI00053999E5 NIAGLVNREDVPIPEKPSSEDPDEIQAWKWSVRKANKINRERHSLRC DVELKLSVARMKM 562
 UPI00053A373F NIAGLVNREDVPIPEKPSSEDPDEIQAWKWSVRKANKINRERHSLRC DVELKLSVARMKM 562
 TR|V4L7M1|V4L7M1_EUTSA NIAGLVNREDVPIPEKPSSEDPPEEQSWKWSVRKANKINRERHSLRC DVELKLSVARMKM 560
 TR|R0HH96|R0HH96_9BRAS NIAGLVNREDVPIPEKPSSEDPPEEQSWKWSVRKANKINRERHSLRC DVELKLSVARMKM 558
 TR|D7LGR7|D7LGR7_ARALL NIAGLVNREDVPIPEKPSSEDPPEEQSWKWSVRKANKINRERHSLRC DVELKLSVARMKM 557
 UPI000859C935 NIAGLVNREDVPIPEKPSSEDPPEEQSWKWSVRKAKKTNRERHSLRC DVELKLSVARMKM 549
 TR|A0A087GS25|A0A087GS25_ARAAL NIAGLVNREDVPIPEKPSSEDPPEEQSWKWSARKANKINRERHSLRC DVELKLSVARMKM 547
 TR|M4FDQ9|M4FDQ9_BRARP NIAGLVNREDVPIPEKPSSEDPPELQSWKWSVRKAKKTNRERHSLRC DVELKLSVARMKM 546
 UPI0006AAEE50 NIAGLVNREDVPIPEKPSSEDPPELQSWKWSVRKAKKTNRERHSLRC DVELKLSVARMKM 546
 *****:*****:*****:*****:*****

SP|O24600|RPOT3_ARATH DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLRCGTLEFAEGRPLGKSGLYWLKIHANLY 620
 TR|A0A1J3GXFO|A0A1J3GXFO_NOCCA DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLRCGTLEFAEGRPLGKQGLYWLKIHANLF 652
 TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLRCGTLEFAEGRPLGKQGLYWLKIHANLF 644
 UPI00053999E5 DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLRCGTLEFAEGRPLGKSGLYWLKIHANLY 622
 UPI00053A373F DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLRCGTLEFAEGRPLGKSGLYWLKIHANLY 622
 TR|V4L7M1|V4L7M1_EUTSA DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLRCGTLEFAEGRPLGKSGLYWLKIHANLY 620
 TR|R0HH96|R0HH96_9BRAS DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLRCGTLEFAEGRPLGKSGLYWLKIHANLY 618
 TR|D7LGR7|D7LGR7_ARALL DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLRCGTLEFAEGRPLGKSGLYWLKIHANLF 617
 UPI000859C935 DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLRCGTLEFAEGRPLGKSGLYWLKIHANLY 609
 TR|A0A087GS25|A0A087GS25_ARAAL DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLRCGTLEFAEGRPLGKSGLYWLKIHANLY 607
 TR|M4FDQ9|M4FDQ9_BRARP DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLRCGTLEFAEGRPLGKSGLYWLKIHANLY 606
 UPI0006AAEE50 DEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLRCGTLEFAEGRPLGKSGLYWLKIHANLY 606
 *****:*****:*****:*****:*****

```

SP|O24600|RPOT3_ARATH      AGGVEKLSHDARLAFVENH LDDIMDSAENPIHGKRWLKAEDPFQCLAACVILTQALKSP 680
TR|A0A1J3GXFO|A0A1J3GXFO_NOCCA AGGVEKLSHDGRILAFVENH LDDIME SAENPVHGKRWLKAEDPFQCLAACVILTQALKSP 712
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA AGGVEKLSHDGRILAFVENH LDDIME SAENPVHGKRWLKAEDPFQCLAACVILTQALKSP 704
UPI00053999E5              AGGVEKLSHDARLAFVENH LDDIMDSAENTIHGKRWLKAEDPFQCLAACVILTQALKSP 682
UPI00053A373F              AGGVEKLSHDARLAFVENH LDDI IDSAENTIHGKRWLKAEDPFQCLAACVILTQALKSP 682
TR|V4L7M1|V4L7M1_EUTSA     AGGVEKLSHDGRILAFVENH LDNIMDSAENPIHGQRWLLKAEDPFQCLAACVILTQALKSP 680
TR|ROHH96|ROHH96_9BRAS     AGGVEKLSHDARLAFVETH LDDVMDSAENPIHGKRWLKAEDPFQCLAACVILTQALKSP 678
TR|D7LGR7|D7LGR7_ARALL    AGGVEKLSHDARLAFVENH LDDIMDSAENPIHGKRWLKAEDPFQCLAACVILTQALKSP 677
UPI000859C935              AGGVEKLSHDGRILAFVENH LDDI IDSAENAIHGKRWLKAEDPFQCLAACVILGQALKSP 669
TR|A0A087GS25|A0A087GS25_ARAAL AGGVEKLSHDGRILAFVENH LDDI IDSAENAIHGKRWLKAEDPFQCLAACVILTQALKSP 667
TR|M4FDQ9|M4FDQ9_BRARP     AGGVEKLSHEGRILSFVENH LDDIMDSAENAIHGRRWLLKAEDPFQCLAACVLAQALKSP 666
UPI0006AAEE50              AGGVEKLSHEGRILSFVENH LDDIMDSAENAIHGRRWLKAEDPFQCLAACVLAQALKSP 666
*****:****:*****:***:****:***:*****:*****:*****:*****:*****:*****:

```

```

SP|O24600|RPOT3_ARATH      SPYSVISHLPIHQDGCSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISRRVHEIMK 740
TR|A0A1J3GXFO|A0A1J3GXFO_NOCCA SPSSVISHLPIHQDGCSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISLRVHEIMK 772
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA SPSSVISHLPIHQDGCSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISLRVHEIMK 764
UPI00053999E5              SPYSVISHLPIHQDGCSCNGLQHYAALGRDSFEAAAVNLVAGDKPADVYSEISRRVHEIMK 742
UPI00053A373F              SPYSVISHLPIHQDGCSCNGLQHYAALGRDSFEAAAVNLVAGDKPADVYSEISRRVHEIMK 742
TR|V4L7M1|V4L7M1_EUTSA     SPYSVISHLPIHQDGCSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISLRVHEIMK 740
TR|ROHH96|ROHH96_9BRAS     SPYSVISHLPIHQDGCSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISRRVHEIMK 738
TR|D7LGR7|D7LGR7_ARALL    SPYSVISHLPIHQDGCSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISRRVHEIMK 737
UPI000859C935              SPYSVISHLPIHQDGCSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISLRVHEIMK 729
TR|A0A087GS25|A0A087GS25_ARAAL SPYSVISHLPIHQDGCSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISLRVHEIMK 727
TR|M4FDQ9|M4FDQ9_BRARP     SPYSVISHLPIHQDGCSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISLRVHEIMK 726
UPI0006AAEE50              SPYSVISHLPIHQDGCSCNGLQHYAALGRDSFEAAAVNLVAGEKPADVYSEISLRVHEIMK 726
** *****:*****:*****:*****:*****:*****:*****:*****:*****:

```

```

SP|O24600|RPOT3_ARATH      KDSSKDPESNPTAALAKILITQVDFKLVKQIVMTSVYGVYVYGAREQIKRRLEEKGVITD 800
TR|A0A1J3GXFO|A0A1J3GXFO_NOCCA KDSNKDPESNPTAALAKILINQVDFKLVKQIVMTSVYGVYVYGAREQIKRRLEEKGVITD 832
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA KDSNKDPESNPTAALAKILINQVDFKLVKQIVMTSVYGVYVYGAREQIKRRLEEKGVITD 824
UPI00053999E5              KDSSKDPESNPTAALAKILITQVDFKLVKQIVMTSVYGVYVYGAREQIKRRLEEKGVITD 802
UPI00053A373F              KDSSKDPESNPTAALAKILITQVDFKLVKQIVMTSVYGVYVYGAREQIKRRLEEKGVITD 802
TR|V4L7M1|V4L7M1_EUTSA     KDSSKDPESNPTAALAKILINQVDFKLVKQIVMTSVYGVYVYGAREQIKRRLEEKGVITD 800
TR|ROHH96|ROHH96_9BRAS     KDSSKDPESNPTAALAKILITQVDFKLVKQIVMTSVYGVYVYGAREQIKRRLEEKGVITD 798
TR|D7LGR7|D7LGR7_ARALL    KDSSKDPESNPTAALAKILITQVDFKLVKQIVMTSVYGVYVYGAREQIKRRLEEKGVITD 797
UPI000859C935              KDSNKDPESNPTAALAKILINQVDFKLVKQIVMTSVYGVYVYGAREQIKRRLEEKGVITD 789
TR|A0A087GS25|A0A087GS25_ARAAL KDSSKDPESNPTAALAKILINQVDFKLVKQIVMTSVYGVYVYGAREQIKRRLEEKGVITD 787
TR|M4FDQ9|M4FDQ9_BRARP     KDSSKDPESNPTAALAKILINQVDFKLVKQIVMTSVYGVYVYGAREQIKRRLEEKGVITD 786
UPI0006AAEE50              KDSSKDPESNPTAALAKILINQVDFKLVKQIVMTSVYGVYVYGAREQIKRRLEEKGVITD 786
***:*****:*****:*****:*****:*****:*****:*****:*****:

```

```

SP|O24600|RPOT3_ARATH      ERMLFAAACYSAKVTLAALGEI FEAARAISWLGDCAKI IASDNHPVRWITPLGLPVPVQP 860
TR|A0A1J3GXFO|A0A1J3GXFO_NOCCA ERMLFAAACYSAKVTLAALGEI FEAARAISWLGDCAKI IASDNHPVRWITPLGLPVPVQP 892
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA ERMLFAAACYSAKVTLAALGEI FEAARAISWLGDCAKI IASDNHPVRWITPLGLPVPVQP 884
UPI00053999E5              ERMLFAAACYSAKVTLAALGEI FEAARAISWLGDCAKI IASDNHPVRWITPLGLPVPVQP 862
UPI00053A373F              ERMLFAAACYSAKVTLAALGEI FEAARAISWLGDCAKI IASDNHPVRWITPLGLPVPVQP 862
TR|V4L7M1|V4L7M1_EUTSA     ERMLFSAACYSAAKVTLAALGEI FEAARAISWLGDCAKI IASDNHPVRWITPLGLPVPVQP 860
TR|ROHH96|ROHH96_9BRAS     ERMLFAAACYSAKVTLAALGEI FEAARAISWLGDCAKI IASDNHPVRWITPLGLPVPVQP 858
TR|D7LGR7|D7LGR7_ARALL    ERMLFAAACYSAKVTLAALGEI FEAARAISWLGDCAKI IASDNHPVRWITPLGLPVPVQP 857
UPI000859C935              ERMLFAAACYSAKVTLAALGEI FEAARAISWLGDCAKI IASDNHPVRWITPLGLPVPVQP 849
TR|A0A087GS25|A0A087GS25_ARAAL ERMLFAAACYSAKVTLAALGEI FEAARAISWLGDCAKI IASDNHPVRWITPLGLPVPVQP 847
TR|M4FDQ9|M4FDQ9_BRARP     ERMLFSAACYSAKVTLAALGEI FEAARAISWLGDCAKI IASDNHPVRWITPLGLPVPVQP 846
UPI0006AAEE50              ERMLFSAACYSAKVTLAALGEI FEAARAISWLGDCAKI IASDNHPVRWITPLGLPVPVQP 846
*****:*****:*****:*****:*****:*****:*****:*****:*****:

```

```

SP|O24600|RPOT3_ARATH      YCRSERHLIRTSLOVLALQREGNTVDVRKQRTA FPPNFVHSLDGTTHMMMTAVACREAGLN 920
TR|A0A1J3GXFO|A0A1J3GXFO_NOCCA YCRSERHLIRTSLOVLALQREGNTVDVRKQRTA FPPNFVHSLDGTTHMMMTAVACREAGLN 952
TR|A0A1J3IJ06|A0A1J3IJ06_NOCCA YCRSERHLIRTSLOVLALQREGNTVDVRKQRTA FPPNFVHSLDGTTHMMMTAVACREAGLN 944
UPI00053999E5              YCRSERHLIRTSLOVLALQREGNTVDVRKQRTA FPPNFVHSLDGTTHMMMTAVACREAGLN 922
UPI00053A373F              YCRSERHLIRTSLOVLALQREGNTVDVRKQRTA FPPNFVHSLDGTTHMMMTAVACREAGLN 922
TR|V4L7M1|V4L7M1_EUTSA     YCRSERHLIRTSLOVLALQREGNTVDVRKQRTA FPPNFVHSLDGTTHMMMTAVACREAGLN 920
TR|ROHH96|ROHH96_9BRAS     YCRSERHLIRTSLOVLALQREGNTVDVRKQRTA FPPNFVHSLDGTTHMMMTAVACREAGLN 918
TR|D7LGR7|D7LGR7_ARALL    YCRSERHLIRTSLOVLALQREGNTVDVRKQRTA FPPNFVHSLDGTTHMMMTAVACREAGLN 917
UPI000859C935              YCRSERHLIRTSLOVLALQREGNTVDVRKQRTA FPPNFVHSLDGTTHMMMTAVACREAGLN 909
TR|A0A087GS25|A0A087GS25_ARAAL YCRNERHLIRTSLOVLALQREGNTVDVRKQRTA FPPNFVHSLDGTTHMMMTAVACREAGLN 907
TR|M4FDQ9|M4FDQ9_BRARP     YCRSERHLIRTSLOVLALQREGNTVDVRKQRTA FPPNFVHSLDGTTHMMMTAVACREAGLN 906
UPI0006AAEE50              YCRSERHLIRTSLOVLALQREGNTVDVRKQRTA FPPNFVHSLDGTTHMMMTAVACREAGLN 906
***:*****:*****:*****:*****:*****:*****:*****:*****:

```

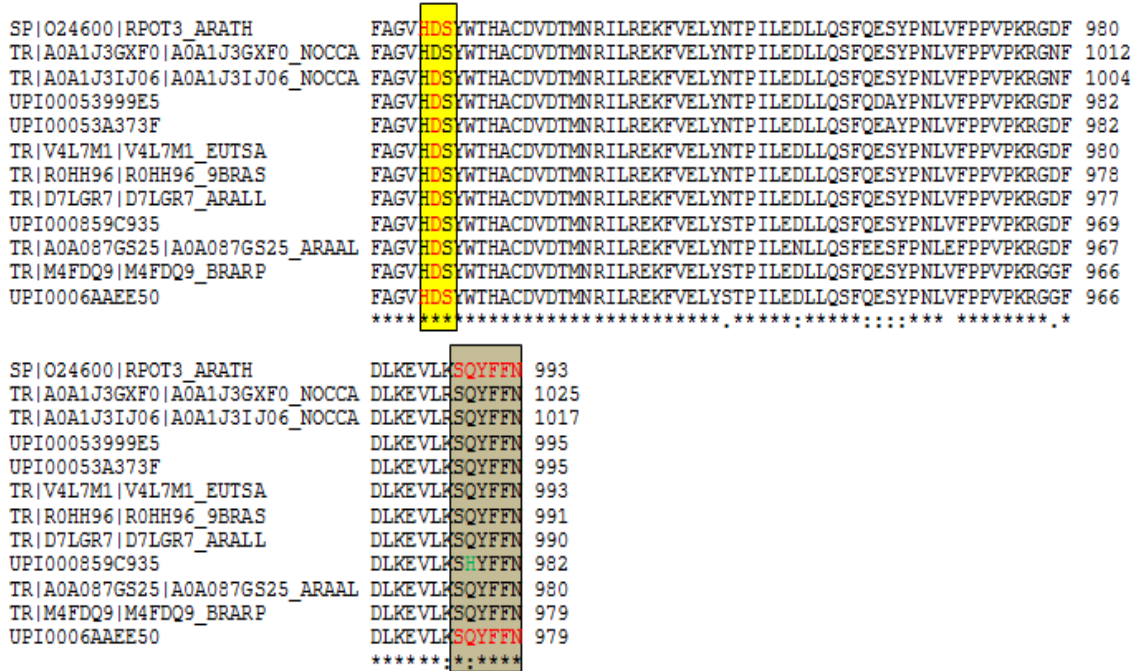


Fig. 3. MSA of various SSU chloroplast RNA polymerases

SP|O24600|RPOT3_ARATH, *Arabidopsis thaliana*
 TR|A0A1J3GXFO|A0A1J3GXFO_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.

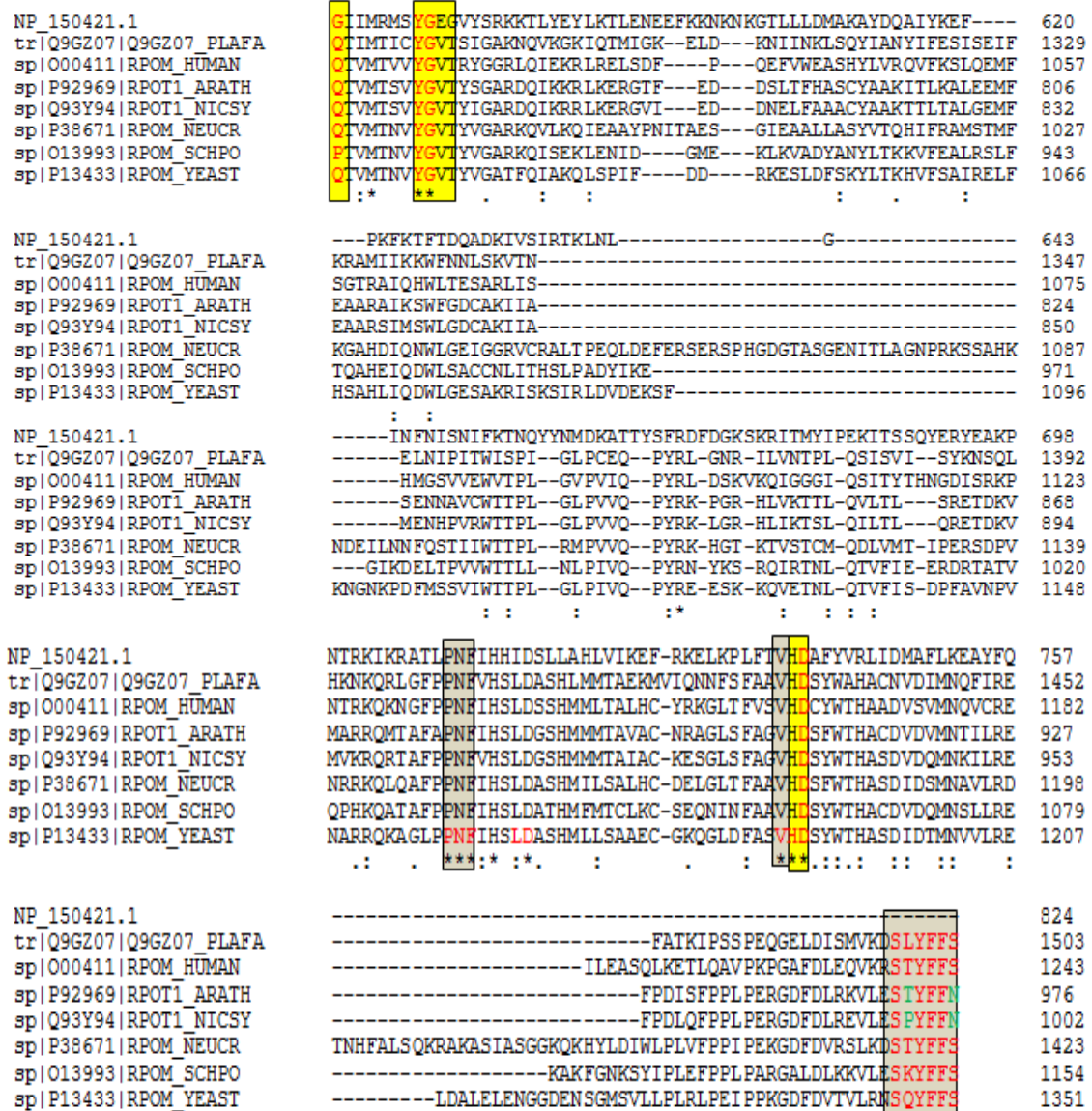


Fig. 4. MSA of mitochondrial RNA polymerases from plants and fungi

*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 glauca*
- 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	AASD9GQLKDLLEGLTALGNVGVKVRKVDMLVKIWNNTGE SFLSIP SAN-TTLDIQEMP	653
sp O13993 RPOM_SCHPO	KASENGQLDELFPKAVSSLGKVSWRINQRLFNVLIRIWNNSGKFLSIP PRE-VKCDMPFPY	646
tr S9Q0Q8 S9Q0Q8_SCHOY	EASHRGHLKRIYNALGALGDVDRINRFTFDVIVKIWNNSGEMLSIP FRN-YEVNLPFPY	660
tr S9X2W4 S9X2W4_SCHCR	EASRRGHLKRVYDALGALGDVSWRINRFTFDVIVKIWNNSGEMLSIP FRN-YEVNLPFPY	659
tr AOA1E3Q3C6 AOA1E3Q3C6_LIPST	EACRRNDLE SVYEGLDVLSAAAWIINTRVFEVLAKVWNTGE EFLIIP TRYEGDINFPLEP	741
tr AOA1E67E4J0 AOA1E67E4J0_9ASCO	AASDRGTLDDQVYEGLDVNLGNTAWTINKRDI FE VILKVVWNTGE EFLIIP RRVEIQPELPPAP	511
tr AOA1E3PUP0 AOA1E3PUP0_9ASCO	EASKRGAMNEVFEGLDVNLGNTAWTINKRDI FE VILKVVWNTGE EFLIIP RRVEIQPELPPAP	716
tr AOA0H5C7R0 AOA0H5C7R0_CYBJA	AASDK--IGKVYEGLDVNLGNTAWTINKRDI FE VILKVVWNTGE EFLIIP RRVEIQPELPPAP	684
tr AOA1E3P5W0 AOA1E3P5W0_WICAO	ASTDR--IDLKVYEGLDVNLGNTAWTINKRDI FE VILKVVWNTGE EFLIIP RRVEIQPELPPAP	693
tr KOKTX3 KOKTX3_WICCF	AASDK--LDKVYDGLNVLGNTAWTINKRDI FE VILKVVWNTGE EFLIIP RRVEIQPELPPAP	720
tr W6MIL2 W6MIL2_9ASCO	AASKRGDLDRVFRGLNVLGNTQWTPNKRILE IVTQVWNSGEMLEIP PAH-SLEKLPDPP	721
tr AOA1E3QFI7 AOA1E3QFI7_9ASCO	AASENGSLASVYKGLTVLGDTPWTVNRRKIYD IVSQVWNTGE SFLDIA GVQ-DELELPPFP	706
tr AOA1D2V948 AOA1D2V948_9ASCO	AASENGDLEGVYDGLNVLGNTAWTINKRDI FE VILKVVWNTGE EFLIIP RRVEIQPELPPAP	746
tr AOA1B7SME0 AOA1B7SME0_9ASCO	TAAIRGKMDTVLQALNNLGS TAWTVNKEVILKVMIQVWNTGE EFLIIP RRVEIQPELPPAP	393
tr Q6CR25 Q6CR25_KLULA	AVSEQGSIDNVYEGLDVNLGNTAWTINKRDI FE VILKVVWNTGE EFLIIP RRVEIQPELPPAP	684
tr WOTGI8 WOTGI8_KLUMA	AVSEQGSIQNVYEGLDVNLGNTAWTINKRDI FE VILKVVWNTGE EFLIIP RRVEIQPELPPAP	683
tr AOA109UM81 AOA109UM81_9SACH	AVSDE--LDNVYKGMNVLGDTFTWTVNKRIMLNIISTIWNNSGEMLEIP PAH-SLEKLPDPP	694
tr G8JMS2 G8JMS2_ERECY	AVTGRGAVNNIYQGLNVLGDTAWTVNKRIFL ILSKIWNNSGE EFLIIP RRVEIQPELPPAP	724
tr Q75BP7 Q75BP7_ASHGO	AVTGRGVVQNVYRGLNVLGDTAWTVNKRIMLNIISTIWNNSGEMLEIP PAH-SLEKLPDPP	717
tr R9XDF6 R9XDF6_ASHAC	AVTGRGAVQNVYRGLNVLGDTAWTVNKRIMLNIISTIWNNSGEMLEIP PAH-SLEKLPDPP	717
tr H2ASJ8 H2ASJ8_KAZAF	AASNVAQLDKVYDGLNVLGDTAWTVNRRIFE IISKVVWNSGEMLEIP PAH-SLEKLPDPP	702
tr J787Y3 J787Y3_KAZNA	AASN3NAIDKVYDGLNVLGDTAWTVNRRIFE IISKVVWNSGEMLEIP PAH-SLEKLPDPP	695
tr GOVDO1 GOVDO1_NAUCC	AVSDADVLPDVYDGLNVLGDTAWTVNRRIFE IISKVVWNSGEMLEIP PAH-SLEKLPDPP	748
tr GOWE72 GOWE72_NAUCC	AVSDKEAIDVYDGLNVLGDTAWTVNRRIFE IISKVVWNSGEMLEIP PAH-SLEKLPDPP	776
tr Q6FLX9 Q6FLX9_CANGA	AVSD3GAIIDKVYDGLNVLGDTAWTVNRRIFE IISKVVWNSGEMLEIP PAH-SLEKLPDPP	746
tr G8B354 G8B354_TETPH	AVSDAGAIIDKVYDGLNVLGDTAWTVNRRIFE IISKVVWNSGEMLEIP PAH-SLEKLPDPP	744
tr AOA0L8RW5 AOA0L8RW5_SACEU	AASENGDIDRVYDGLNVLGDTFTWTVNRRKVFVVSQVWNSGEMLEIP PAH-SLEKLPDPP	771
tr AOA0L8VRU3 AOA0L8VRU3_9SACH	AASDNGDIDRVYDGLNVLGDTFTWTVNRRKVFVVSQVWNSGEMLEIP PAH-SLEKLPDPP	770
tr J8P588 J8P588_SACAR	AASENGDIDRVYDGLNVLGDTFTWTVNRRKVFVVSQVWNSGEMLEIP PAH-SLEKLPDPP	770
tr AOA0C7MY71 AOA0C7MY71_9SACH	AVSD3GAIIDKVYDGLNVLGDTFTWTVNRRKVFVVSQVWNSGEMLEIP PAH-SLEKLPDPP	716
tr C5DNF3 C5DNF3_LACTC	AVSDAGAIIDKVYDGLNVLGDTFTWTVNRRKVFVVSQVWNSGEMLEIP PAH-SLEKLPDPP	711
tr C5DX79 C5DX79_2YGRC	AVSDATAINTVYDGLNVLGDTFTWTVNRRKVFVVSQVWNSGEMLEIP PAH-SLEKLPDPP	732
tr G8ZRO0 G8ZRO0_TORDC	AVSNAGAIDTVYQGLNVLGDTAWTVNRRVDFVMSEVWNSGEMLEIP PAH-SLEKLPDPP	742
tr AOA1E4RQF7 AOA1E4RQF7_9ASCO	AASN--RISGVYDGLNVLGDTFTWTVNRRKVFVVSQVWNSGEMLEIP PAH-SLEKLPDPP	737
tr AOA0L0P4K6 AOA0L0P4K6_9ASCO	AASDAHNLDEVYRGLNVLGDTAWTVNRRKVFVVSQVWNSGEMLEIP PAH-SLEKLPDPP	725
tr AOA1A0HGT7 AOA1A0HGT7_9ASCO	AASD9GRILDGVYAGLDVNLGNTAWTVNRRVDFVMSEVWNSGEMLEIP PAH-SLEKLPDPP	728
tr C4Y8E3 C4Y8E3_CLAL4	AASEANNLDDVYRGLNVLGHTPWTINAKVLEVISQVWNTGE AFLDIP FVV-DEPELPPFL	731
tr G3B4C1 G3B4C1_CANTC	AASQGRDLDRVYDGLNVLGNTFTWTVNRRKVFVVSQVWNSGEMLEIP PAH-SLEKLPDPP	703
tr A3LX46 A3LX46_PICST	KAADLGNLNEVYDGLNVLGKTFWTVNRRVFE IITRYWNSGEMLEIP PAH-SLEKLPDPP	677
tr AOA1E4SMT6 AOA1E4SMT6_9ASCO	AASEMGNLDEIYQGLNVLGNTAWTVNRRVDFVMSEVWNSGEMLEIP PAH-SLEKLPDPP	678
tr A5DN82 A5DN82_PICGU	AASDMGNLDDQVYEGLDVNLGTECWTINHEVDFVISHYWNNSGEMLEIP PAH-SLEKLPDPP	723
tr B5RTF6 B5RTF6_DEBHA	AASDLDNLE-IYDGLNVLGDTAWTVNRRVFN IISKFWNTGE EFLIIP RRVEIQPELPPAP	732
tr G3AEY0 G3AEY0_SPAFN	TAARNGNLDQVYAGLDVNLGNTAWTVNRRVFN IISKFWNTGE EFLIIP RRVEIQPELPPAP	713
tr G8B7X1 G8B7X1_CANPC	AAAKRGNLKEVFDGLNVLGDTAWTINKRDI FE VILKVVWNTGE EFLIIP RRVEIQPELPPAP	718
tr H8X1L6 H8X1L6_CANO9	AAAKRGNLQEVFDGLNVLGDTAWTINKRDI FE VILKVVWNTGE EFLIIP RRVEIQPELPPAP	717
tr B9W6L5 B9W6L5_CANDC	AAAREGNLKAFFEGLDVNLGKTAWTVNRRVFN IISKFWNTGE EFLIIP RRVEIQPELPPAP	726
tr C4YFJ1 C4YFJ1_CANAW	AAAREGNLITGVYEGLDVNLGNTAWTVNRRVFN IISKFWNTGE EFLIIP RRVEIQPELPPAP	706
tr CSME71 CSME71_CANTI	AASKRGNLDQVFDGLNVLGNTAWTVNRRVFN IISKFWNTGE EFLIIP RRVEIQPELPPAP	721
tr M3IK19 M3IK19_CANMX	AASKRGNLDEVFRGLNVLGSDFTWTVNRRVFN IISKFWNTGE EFLIIP RRVEIQPELPPAP	723

tr B6K333 B6K333_SCHJY	FVNDHLQDI FNSVDKFLDGE-QFWTKADDF	QALAVCFEIA DAIRSGDFSSF ISHVFP	827
tr 013993 RFOM_SCHPO	FVDDNMQEV FDSADR PLDGN-KWWSKADDF	QALAA CFEI AEAVRSGDHESY IISHIPL	820
tr 39Q0Q8 39Q0Q8_SCHYO	FVDDNIAEL FDSYDH FLEGR-KWWSAEDDF	QALAA IAEI ARAIRSGNFESY VCHVFP	834
tr 39X2W4 39X2W4_SCHCR	FVDDNIEEV FDSYDR PLDGR-KWWT SAEDDF	QALAA IAEI ARAIRSGNFESY VCHVFP	833
tr A0A1E3Q3C6 A0A1E3Q3C6_LIPST	YADDKIQEI FDSADC PLDGR-RWWWQASDF	QCLAA CFELA DALRSF DPFYF RSRLP	915
tr A0A1E7E4J0 A0A1E7E4J0_9ASCO	FADNMEKI LDSANS PMEQQ-KWWQEAESDF	QALAT CIEIRNAMLLEDP3KY KCRLP	685
tr A0A1E3PUP0 A0A1E3PUP0_9ASCO	FADNHKIDI YDSEAK PLDGG-KWMTADDF	QALAV CMELS KAYSMD DPTKF VSRLP	890
tr A0A0H5C7R0 A0A0H5C7R0_CYBJA	YAEENVNIMKTAAD PLGFQ-DWWT KADKFP	QALAT IFELA DALKLP DPTKY VSHQFP	859
tr A0A1E3P5W0 A0A1E3P5W0_WICAO	FADNLENI KRTAED PMANK-EFWT KADKFP	QALAT IYDLA DAYKLE DPTKH ISHQFP	867
tr K0KTX3 K0KTX3_WICCF	FVNDNIELVKRAED PMANG-EWWT KADKFP	QALAT IIDLA DALKLP DPTKH ISHQFP	894
tr W6MIL2 W6MIL2_9ASCO	FTEKHLEDI RDSARD PLGGG-RWWWKADKFP	QALAT CFELE AAFNLE DPTQF ISHQFP	899
tr A0A1E3QPI7 A0A1E3QPI7_9ASCO	FADAHVAE IMESARD PLHGA-AWKKADKFP	QALST IFELS EALQMA DPTQV VSHQFP	880
tr A0A1D2V948 A0A1D2V948_9ASCO	FTENNYENI HKSAEN PQAAD-AWKKADKFP	QALAT IFELS QALKLP DPTKF ISHQFP	920
tr A0A1B7SME0 A0A1B7SME0_9ASCO	YVDDHLDEI FESARD PLGGG-RWWV KADKFP	QFLAS AMELE QAFRLP DPTKF ISHQFP	867
tr Q6CR25 Q6CR25_KLULA	FADT HLKEI RESAEH PLDGT-RWWT KADKFP	QFLAT CIELNEALKLDNPNEDF ISHQFP	858
tr W0TGI8 W0TGI8_KLUMA	FANAHFDDI KDSAEN PLGGK-RWWT KADKFP	QFLAT CIELNEAMKLDNPNEDF ISHQFP	867
tr A0A109UWS1 A0A109UWS1_9SACH	FTEFNLEHV KESAEN PLDGR-GWKKADKFP	QCLAT CLELNEAYKLENPEDF VSHQFP	858
tr G8JMS2 G8JMS2_ERECY	FTEE HLEDI KDSAEN PLHGN-GWKKADKFP	QCLAT CMELR DAYKLENPEDF ISHQFP	898
tr Q75BP7 Q75BP7_ASHGO	FTEAHLDEI KDSAEN PLNGK-GWKKADKFP	QCLAT CMEINNAVYKLSNPEDV VSHQFP	895
tr R9XDF6 R9XDF6_ASHAC	FTEAHLDEI KESAEN PLNGK-GWKKADKFP	QCLAT CMEINNAVYKLSNPEDV VSHQFP	895
tr H2A3J8 H2A3J8_KAZAF	FIDHLEEI KDSAEN PLNGK YLWQKADKFP	QALAT CIELNEALKLENPEDF ISHQFP	877
tr J797Y3 J797Y3_KAZNA	FIE3QLDEV KDSAE DPLNGK-GWKKADKFP	QALAT CMEINNAKLDNPNEDF VSHQFP	869
tr G0VDO1 G0VDO1_NAUCC	FVEDHLQEI KESAEN PLTTG-KWKKADKFP	QCLAT CIELT EALKLDNPEEF ISHQFP	922
tr G0WE72 G0WE72_NAUCC	FIEDHLEI DVKDTAEN PLDGG-GWKKADKFP	QCLAT CIELNEALKLDNPNEDF ISHQFP	950
tr Q6FLX9 Q6FLX9_CANGA	FTECHIEDI KDSAEN PLNGK-GWKKADKFP	QALAT CMEINNAKLDNPNEDV ISHQFP	920
tr G8B554 G8B554_TETPH	FAEENMENI KDSAEN PLNGK-GWKKADKFP	QALST CIELNEAFLKNPEDF ISHQFP	918
tr A0A0L8RWK5 A0A0L8RWK5_SACEU	FTE3SHLKI KDSAEN PLTGD-RWWT TADKFP	QALAT CFELENEVMKLDNPNEDF VSHQFP	945
tr A0A0L8VRU3 A0A0L8VRU3_9SACH	FTE3SHLQDI KDSAEN PLTGD-RWWT TADKFP	QALAT CFELENEVMKLDNPNEDF ISHQFP	944
tr J8P558 J8P558_SACAR	FTE3SHLEI KDSAEN PLTGG-GWKKADKFP	QALAT CFELENEVMKLDNPNEDF ISHQFP	944
tr A0A0C7M7Y1 A0A0C7M7Y1_9SACH	FTE3SHLEI KDSAEN PLDGG-GWKKADKFP	QALAT CMEINNAKLDNPNEDF ISHVFP	890
tr C5DNF3 C5DNF3_LACTC	FAEDHIDDI RDSAEH PLDGG-CWKKADKFP	QALAT CMEINNAKLDNPNEDF VSHQFP	885
tr C5DX79 C5DX79_2YGRK	FAEDHIDDI KQSAQD PLKED-AWKKADKFP	QALAT CMEINNAKLDNPNEDF ISHQFP	906
tr G82R00 G82R00_TORDC	YAEAHLEI KQSAQD PLDGG-GWKKADKFP	QALAT CIEINEAYKLPNPEDF ISHVFP	916
tr A0A1E4RQF7 A0A1E4RQF7_9ASCO	FANKNIENI KAVAKD PYQ-N-CWVVGKDFP	QILGI CYELAEAYLTDPTKF VSHVFP	910
tr A0A0L0P4K6 A0A0L0P4K6_9ASCO	FVEENMDKV FQTARD PLGED-RWVI KGDKFP	QVLSV CFELNEAYKLE DPTKY VSYVFP	922
tr A0A1A0HG7 A0A1A0HG7_9ASCO	FVDDNLENI CKVAEN PIANG-EWWS KADKFP	QVLSV CFELNEAHLKLD DPTKY VSYVFP	902
tr C4Y8E3 C4Y8E3_CLAL4	FVDDNLEHV FESARD PTGGG-KWWT KGEKFP	QVLSV CFELNEAHLKLD DPTKY VSYVFP	905
tr G3B4C1 G3B4C1_CANTC	FTEETHLEI RESAET PFDDG-AWWT KGEKFP	QVLSV CFELNEAYKLE DPTKY VSHVFP	877
tr A3LX46 A3LX46_PIC9T	FVDDNLEHV FESARN PYDGD-AWKKADKFP	QALGV CFELE EAYKLENPTQV VSHIPI	851
tr A0A1E43MT6 A0A1E43MT6_9ASCO	FVNNENLENI PQSAQN PYDQ-AWWT KGEKFP	QVLSV CFELNEAYKMS DPTKY VSHIPI	852
tr A5DN82 A5DN82_PICGU	FVNNLEHNI IQSAKN PLDGD-AWKKADKFP	QVLSV CFELE KRAYLED DPTKH ISHLPI	897
tr B5RTF6 B5RTF6_DEBHA	FVDDNLENI LECARN PINGS-GWKKADKFP	QVLSV CFELNEAFLQD DPTKY VSHIPI	906
tr G3AEY0 G3AEY0_SAPFN	FVDDNLQNV FESAKN PY3ED-AWKKADKFP	QALGV CFELE RRAYKLP DPTKY VSHMVA	887
tr G8B7X1 G8B7X1_CANFC	FATDHLQEA IKSAED PLN-H-KWWT KAEKFP	QALSV CFELA EAYKLP DPTKY VSYLPI	891
tr H8X1L6 H8X1L6_CAN09	FATDHLQEA IKSAED PLE-Y-KWWT KAEKFP	QALSV CFELA EAYKLP DPTKY VSYLPI	890
tr B9W6L5 B9W6L5_CANDC	FVDDNLEKV FASAAD PLASN-AWVQ KAEKFP	QALSV CFELA EAYKLD DPTKY VSHLPI	900
tr C4YFJ1 C4YFJ1_CANAM	FVNNENLEHV LE3AAD PFATD-AWVQ KAEKFP	QALSV CFELA EAYKLD DPTKY VSHLPI	880
tr C5ME71 C5ME71_CANTT	FIDDNTIENV IASAKD PYAPD-AWVQRAEKFP	QALSV CFELGEAYQLD DPTKY VSHLPI	895
tr M3IK19 M3IK19_CANMX	FVDD SMDKV IE3AAD PL3AED-AWKKADKFP	QALSV CFELA EAYKLD DPTKY VSHLPI	897

tr B6K333 B6K333_SCHJY	tr STCNGQLQYAAALGDIEGAGQVNL	tr FSPNRPNVYAAVAARVISILKKE--AAGDPM	885
tr 013993 RFOM_SCHPO	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPSDHP SDVVE AVAE IVRGFLKLD--AEAGDE	878
tr 39Q0Q8 39Q0Q8_SCHYO	tr STCNGQLQYAAALGRDPDGAHEVNL	tr SPNDRP KDVDY DAVAKI VISRLEQE--S	892
tr 39X2W4 39X2W4_SCHCR	tr STCNGQLQYAAALGRDPDGAHEVNL	tr SPNDRP KDVDY DAVAKI VISKLEHE--SM	891
tr A0A1E3Q3C6 A0A1E3Q3C6_LIPST	tr STCNGQLQYAAALGDIEGAGQVNL	tr EPGDKF QDIYI HVASRVHDYVQKD--A	892
tr A0A1E7E4J0 A0A1E7E4J0_9ASCO	tr STCNGQLQYAAALGDIEGAGQVNL	tr VPS8KF QDVIYS RVLE IVRVRVEED--	743
tr A0A1E3PUP0 A0A1E3PUP0_9ASCO	tr STCNGQLQYAAALGDIEGAGQVNL	tr IF8DKF QDIYS EVANI VRARINND--F	948
tr A0A0H5C7R0 A0A0H5C7R0_CYBJA	tr STCNGQLQYAAALGDIEGAGQVNL	tr FPAERF QDVYI FVANLVKERLKL--A	917
tr A0A1E3P5W0 A0A1E3P5W0_WICAO	tr STCNGQLQYAAALGDIEGAGQVNL	tr VPADRQ QDVYI HVAGLVTKRLEAA--A	925
tr K0KTX3 K0KTX3_WICCF	tr STCNGQLQYAAALGDIEGAGQVNL	tr AP8DKF QDVYI HVAKLVRAKLEEA--	952
tr W6MIL2 W6MIL2_9ASCO	tr STCNGQLQYAAALGDIEGAGQVNL	tr FPADRQ KDVIY HVAGVQVTERLKE--	857
tr A0A1E3QPI7 A0A1E3QPI7_9ASCO	tr STCNGQLQYAAALGDIEGAGQVNL	tr SFLERF QDVYI FVAELVVKRLEAA--D	935
tr A0A1D2V948 A0A1D2V948_9ASCO	tr STCNGQLQYAAALGDIEGAGQVNL	tr SFS8KF QDVYI YVAGLVQARVDED--	978
tr A0A1B7SME0 A0A1B7SME0_9ASCO	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPA8KF SDVYI HVAKLVKRSVQED--	625
tr Q6CR25 Q6CR25_KLULA	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI FVAKLVTKRLENS--	916
tr W0TGI8 W0TGI8_KLUMA	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI FVAKLVBERLKA--	815
tr A0A109UWS1 A0A109UWS1_9SACH	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI HVAKLVTERLKA--	926
tr G8JMS2 G8JMS2_ERECY	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI FVAKLVTERLKA--	956
tr Q75BP7 Q75BP7_ASHGO	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI FVAKLVTERLKA--	953
tr R9XDF6 R9XDF6_ASHAC	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI FVAKLVTERLKA--	953
tr H2A3J8 H2A3J8_KAZAF	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI HVAKLVQKRLRQDNDH	937
tr J797Y3 J797Y3_KAZNA	tr STCNGQLQYAAALGDIEGAGQVNL	tr IP8AMP KDVIYS HVANI VSRVLQVL--	928
tr G0VDO1 G0VDO1_NAUCC	tr STCNGQLQYAAALGDIEGAGQVNL	tr IP8DRF QDVYI HVAKLVTRKLEKA--	980
tr G0WE72 G0WE72_NAUCC	tr STCNGQLQYAAALGDIEGAGQVNL	tr IP8NERF QDVYI HVAKLVTRKLEKA--	1008
tr Q6FLX9 Q6FLX9_CANGA	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI FVAKLVLEARLEKA--	978
tr G8B554 G8B554_TETPH	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI HVAKLVTRKLEKA--	976
tr A0A0L8RWK5 A0A0L8RWK5_SACEU	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPNDRF QDVYI HVAKLVQKRLEIA--	1003
tr A0A0L8VRU3 A0A0L8VRU3_9SACH	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI HVAKLVQKRLEIA--	1002
tr J8P558 J8P558_SACAR	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI HVAKLVQKRLEIA--	1002
tr A0A0C7M7Y1 A0A0C7M7Y1_9SACH	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI HVAKLVVARLEAA--	948
tr C5DNF3 C5DNF3_LACTC	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI HVAKLVVARLEKA--	943
tr C5DX79 C5DX79_2YGRK	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPNDRF QDVYI HVAKLVVARLEIA--	964
tr G82R00 G82R00_TORDC	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPNDRF QDVYI HVAKLVQARLEAA--	974
tr A0A1E4RQF7 A0A1E4RQF7_9ASCO	tr STCNGQLQYAAALGDIEGAGQVNL	tr KPADRF QDVYI YVAGLVIKRVKAD--	968
tr A0A0L0P4K6 A0A0L0P4K6_9ASCO	tr STCNGQLQYAAALGDIEGAGQVNL	tr LPADRQ QDVYI HVAGLVKRLDAE--	957
tr A0A1A0HG7 A0A1A0HG7_9ASCO	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI FVA8LVKRLVE--	963
tr C4Y8E3 C4Y8E3_CLAL4	tr STCNGQLQYAAALGDIEGAGQVNL	tr VPADRQ QDVYI FVAGLVKRLQAE--	960
tr G3B4C1 G3B4C1_CANTC	tr STCNGQLQYAAALGDIEGAGQVNL	tr VPADRQ QDVYI SVASIVQQRVAD--	935
tr A3LX46 A3LX46_PIC9T	tr STCNGQLQYAAALGDIEGAGQVNL	tr LPADRQ QDVYI FVAGLVQKRIDA--	909
tr A0A1E43MT6 A0A1E43MT6_9ASCO	tr STCNGQLQYAAALGDIEGAGQVNL	tr LPADRQ QDVYI FVAGLVQKRIDA--	910
tr A5DN82 A5DN82_PICGU	tr STCNGQLQYAAALGDIEGAGQVNL	tr IPADRQ QDVYI FVAGLVQKRIDA--	955
tr B5RTF6 B5RTF6_DEBHA	tr STCNGQLQYAAALGDIEGAGQVNL	tr IPADRQ QDVYI FVAGLVQKRIDA--	964
tr G3AEY0 G3AEY0_SAPFN	tr STCNGQLQYAAALGDIEGAGQVNL	tr ARADRQ QDVYI YVAGLVRELEKD--	945
tr G8B7X1 G8B7X1_CANFC	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI FVSKLVQNRVAD--	949
tr H8X1L6 H8X1L6_CAN09	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI FVSKLVQNRVAD--	948
tr B9W6L5 B9W6L5_CANDC	tr STCNGQLQYAAALGDIEGAGQVNL	tr KPADRF QDVYI YVAKLVVARVAD--	949
tr C4YFJ1 C4YFJ1_CANAM	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI YVAKLVTRKAD--	938
tr C5ME71 C5ME71_CANTT	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI YVAKLVTRKAD--	938
tr M3IK19 M3IK19_CANMX	tr STCNGQLQYAAALGDIEGAGQVNL	tr WPS8RF QDVYI YVAGLVKRVAD--	955


```

tr|B6K333|B6K333_SCHJY
sp|O13993|RPOM_SCHFO
tr|S9Q0Q8|S9Q0Q8_SCHOY
tr|S9X2W4|S9X2W4_SCHCR
tr|A0A1E3Q3C6|A0A1E3Q3C6_LIPST
tr|A0A1E7E4J0|A0A1E7E4J0_9ASCO
tr|A0A1E3PUP0|A0A1E3PUP0_9ASCO
tr|A0A0H5C7R0|A0A0H5C7R0_CYBJA
tr|A0A1E3P5W0|A0A1E3P5W0_WICAO
tr|K0KTX3|K0KTX3_WICCF
tr|W6MIL2|W6MIL2_9ASCO
tr|A0A1E3QPI7|A0A1E3QPI7_9ASCO
tr|A0A1D2V948|A0A1D2V948_9ASCO
tr|A0A1B73ME0|A0A1B73ME0_9ASCO
tr|Q6CRZ5|Q6CRZ5_KLULA
tr|W0TG18|W0TG18_KLUMA
tr|A0A0L8VRU3|A0A0L8VRU3_9SACH
tr|G8JMS2|G8JMS2_ERECY
tr|Q75BP7|Q75BP7_ASHGO
tr|R9XDF6|R9XDF6_ASHAC
tr|H2A3J8|H2A3J8_KAZAF
tr|J7S7Y3|J7S7Y3_KAZNA
tr|G0VD01|G0VD01_NAUCC
tr|G0WE72|G0WE72_NAUCC
tr|Q6FLX9|Q6FLX9_CANGA
tr|G8B354|G8B354_TETPH
tr|A0A0L8RWK5|A0A0L8RWK5_SACEU
tr|A0A0L8VRU3|A0A0L8VRU3_9SACH
tr|J8P582|J8P582_SACAR
tr|A0A0C7MY71|A0A0C7MY71_9SACH
tr|C5DNF3|C5DNF3_LACTC
tr|C5DX79|C5DX79_ZYGRC
tr|G8ZRO0|G8ZRO0_TORDC
tr|A0A1E4RQF7|A0A1E4RQF7_9ASCO
tr|A0A0L0P4K6|A0A0L0P4K6_9ASCO
tr|A0A1A0HG77|A0A1A0HG77_9ASCO
tr|C4Y8E3|C4Y8E3_CLA14
tr|G3B4C1|G3B4C1_CANTC
tr|A3LX46|A3LX46_PIC3T
tr|A0A1E4SMT6|A0A1E4SMT6_9ASCO
tr|A5DN82|A5DN82_PICGU
tr|B5RTF6|B5RTF6_DEBHA
tr|G3AEY0|G3AEY0_SFAPN
tr|G8B7X1|G8B7X1_CANFC
tr|H8X1L6|H8X1L6_CAN09
tr|B9W6L5|B9W6L5_CANDC
tr|C4YFJ1|C4YFJ1_CANAW
tr|C5ME71|C5ME71_CANTT
tr|M3IK19|M3IK19_CANMX

```

```

tr|B6K333|B6K333_SCHJY
sp|O13993|RPOM_SCHFO
tr|S9Q0Q8|S9Q0Q8_SCHOY
tr|S9X2W4|S9X2W4_SCHCR
tr|A0A1E3Q3C6|A0A1E3Q3C6_LIPST
tr|A0A1E7E4J0|A0A1E7E4J0_9ASCO
tr|A0A1E3PUP0|A0A1E3PUP0_9ASCO
tr|A0A0H5C7R0|A0A0H5C7R0_CYBJA
tr|A0A1E3P5W0|A0A1E3P5W0_WICAO
tr|K0KTX3|K0KTX3_WICCF
tr|W6MIL2|W6MIL2_9ASCO
tr|A0A1E3QPI7|A0A1E3QPI7_9ASCO
tr|A0A1D2V948|A0A1D2V948_9ASCO
tr|A0A1B73ME0|A0A1B73ME0_9ASCO
tr|Q6CRZ5|Q6CRZ5_KLULA
tr|W0TG18|W0TG18_KLUMA
tr|A0A0L8VRU3|A0A0L8VRU3_9SACH
tr|G8JMS2|G8JMS2_ERECY
tr|Q75BP7|Q75BP7_ASHGO
tr|R9XDF6|R9XDF6_ASHAC
tr|H2A3J8|H2A3J8_KAZAF
tr|J7S7Y3|J7S7Y3_KAZNA
tr|G0VD01|G0VD01_NAUCC
tr|G0WE72|G0WE72_NAUCC
tr|Q6FLX9|Q6FLX9_CANGA
tr|G8B354|G8B354_TETPH
tr|A0A0L8RWK5|A0A0L8RWK5_SACEU
tr|A0A0L8VRU3|A0A0L8VRU3_9SACH
tr|J8P582|J8P582_SACAR
tr|A0A0C7MY71|A0A0C7MY71_9SACH
tr|C5DNF3|C5DNF3_LACTC
tr|C5DX79|C5DX79_ZYGRC
tr|G8ZRO0|G8ZRO0_TORDC
tr|A0A1E4RQF7|A0A1E4RQF7_9ASCO
tr|A0A0L0P4K6|A0A0L0P4K6_9ASCO
tr|A0A1A0HG77|A0A1A0HG77_9ASCO
tr|C4Y8E3|C4Y8E3_CLA14
tr|G3B4C1|G3B4C1_CANTC
tr|A3LX46|A3LX46_PIC3T
tr|A0A1E4SMT6|A0A1E4SMT6_9ASCO
tr|A5DN82|A5DN82_PICGU
tr|B5RTF6|B5RTF6_DEBHA
tr|G3AEY0|G3AEY0_SFAPN
tr|G8B7X1|G8B7X1_CANFC
tr|H8X1L6|H8X1L6_CAN09
tr|B9W6L5|B9W6L5_CANDC
tr|C4YFJ1|C4YFJ1_CANAW
tr|C5ME71|C5ME71_CANTT
tr|M3IK19|M3IK19_CANMX

```

tr B6K333 B6K333_SCHJY	LPVVQYRQKFFPSKQITINLQSVYLEDPDSDMAFVDPRKQTAV	PPNF IHS LDA THIMTCL	1084
tr O13993 RPM_SCHRO	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1047
tr S9Q008 S9Q008_SCHYO	FP IIVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNFVHSLDATHIMTCL	1061
tr S9X294 S9X294_SCHCR	FP IIVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNFVHSLDATHIMTCL	1060
tr A0A1E3Q3C6 A0A1E3Q3C6_LIPST	LPVVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1143
tr A0A1E7E4J0 A0A1E7E4J0_SASCO	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	919
tr A0A1E2PUP0 A0A1E2PUP0_SASCO	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1122
tr A0A0H5C7R0 A0A0H5C7R0_CYBJA	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1090
tr A0A1E3P5W0 A0A1E3P5W0_WICAO	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1098
tr K0KTX3 K0KTX3_WICCF	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1125
tr W6MIL2 W6MIL2_SASCO	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNFVHSLDATHIMTCL	1129
tr A0A1E3QPI7 A0A1E3QPI7_SASCO	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1100
tr A0A1D2V948 A0A1D2V948_SASCO	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1152
tr A0A1B7SME0 A0A1B7SME0_SASCO	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1100
tr Q6CR25 Q6CR25_KLULA	LPVVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	793
tr W0TGI8 W0TGI8_KLUMA	LPVVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1089
tr A0A109UWS1 A0A109UWS1_S9ACH	LPVVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1088
tr G8JMS2 G8JMS2_ERECY	LPVVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1099
tr Q75BP7 Q75BP7_ASHGO	LPVVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1129
tr R9XDF6 R9XDF6_ASHAC	LPVVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1124
tr H2ASJ8 H2ASJ8_KAZAF	LPVVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1110
tr J75TY3 J75TY3_KAZNA	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1104
tr G0VD01 G0VD01_NAUCC	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1153
tr G0WE72 G0WE72_NAUDC	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1181
tr Q6FLX9 Q6FLX9_CANGA	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1151
tr G8BS54 G8BS54_TETPH	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1149
tr A0A0L8RKW5 A0A0L8RKW5_SACEU	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1176
tr A0A0L8VRU3 A0A0L8VRU3_S9ACH	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1175
tr J8PP58 J8PP58_SACAR	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1175
tr A0A0C7MY71 A0A0C7MY71_S9ACH	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1121
tr C5DNP3 C5DNP3_LACTC	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1116
tr C5DX79 C5DX79_ZYGRC	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1137
tr G8ZR00 G8ZR00_TORDC	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1143
tr A0A1E4RQF7 A0A1E4RQF7_SASCO	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNFVHSLDATHIMTCL	1147
tr A0A0L0P4K6 A0A0L0P4K6_SASCO	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNFVHSLDATHIMTCL	1132
tr A0A1A0HGT7 A0A1A0HGT7_SASCO	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1126
tr C4YB23 C4YB23_CLAL4	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1139
tr G3B4C1 G3B4C1_CANTC	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNFVHSLDATHIMTCL	1108
tr A3LX46 A3LX46_FICST	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNFVHSLDATHIMTCL	1081
tr A0A1E4SMT6 A0A1E4SMT6_SASCO	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNFVHSLDATHIMTCL	1068
tr A5DM52 A5DM52_FICCG	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNFVHSLDATHIMTCL	1180
tr B5RTF6 B5RTF6_DBBHA	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNFVHSLDATHIMTCL	1139
tr G3AEY0 G3AEY0_SPAFN	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNFVHSLDATHIMTCL	1116
tr G8B7X1 G8B7X1_CANPC	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1120
tr H8X1L6 H8X1L6_CAN09	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1120
tr B9W6L5 B9W6L5_CANDC	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1128
tr C4YFJ1 C4YFJ1_CANAW	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1108
tr C5ME71 C5ME71_CANTT	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNFVHSLDATHIMTCL	1121
tr M3IK19 M3IK19_CANMX	LP IVQYRQKFFPSKQIRTNLQSVFIEERRDRTAVVQPHQATA	PPNF IHS LDA THIMTCL	1126
tr B6K333 B6K333_SCHJY	KCKKRNITVFVVVVVVVVVVAASVAELNGLLREAFVELHSDI	LDGLKRFEDRYKXNVYH	1114
tr O13993 RPM_SCHRO	KCEEQNINFAVVRVVRVVAACVDVQMNILRLAEVLLHNSN	IMERLKEFERRYKQFLV	1107
tr S9Q008 S9Q008_SCHYO	KAKEDAITFABVVRVVRVVAACVDVQMNILRLAEVLLHNSN	IMERLKEFERRYKQFLV	1121
tr S9X294 S9X294_SCHCR	KAKAANITFABVVRVVRVVAACVDVQMNILRLAEVLLHNSN	IMERLKEFERRYKXNVYH	1120
tr A0A1E3Q3C6 A0A1E3Q3C6_LIPST	ACCRNGLTFAVVRVVRVVAADI DRMAIILRLDAF IKLHECD	VAKLKEEFERRYKQYVQ	1203
tr A0A1E7E4J0 A0A1E7E4J0_SASCO	CCGREGISFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	979
tr A0A1E3PUP0 A0A1E3PUP0_SASCO	ECGKKNLSFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1182
tr A0A0H5C7R0 A0A0H5C7R0_CYBJA	ACSKKEGLFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1180
tr A0A1E3P5W0 A0A1E3P5W0_WICAO	SCRKEGLFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1188
tr K0KTX3 K0KTX3_WICCF	CSCKKLEFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1185
tr W6MIL2 W6MIL2_SASCO	ACSEKNISFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1189
tr A0A1E3QPI7 A0A1E3QPI7_SASCO	KCRADNLFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1180
tr A0A1D2V948 A0A1D2V948_SASCO	ECGRKGLFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1212
tr A0A1B7SME0 A0A1B7SME0_SASCO	KCSEAGLFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	853
tr Q6CR25 Q6CR25_KLULA	QCGKAGLDFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1149
tr W0TGI8 W0TGI8_KLUMA	QCGKAGLDFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1148
tr A0A109UWS1 A0A109UWS1_S9ACH	A9AKHGLDFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1189
tr G8JMS2 G8JMS2_ERECY	ECG9GQLFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1189
tr Q75BP7 Q75BP7_ASHGO	ECGRGLGLFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1184
tr R9XDF6 R9XDF6_ASHAC	ECGKAGLDFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1184
tr H2ASJ8 H2ASJ8_KAZAF	ECXKGLGLFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1170
tr J75TY3 J75TY3_KAZNA	KCAEEGLDFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1161
tr G0VD01 G0VD01_NAUCC	QCRANGLEFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1213
tr G0WE72 G0WE72_NAUDC	KCKENGLFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1241
tr Q6FLX9 Q6FLX9_CANGA	GCRRNGLDFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1211
tr G8BS54 G8BS54_TETPH	ACCKEGLDFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1209
tr A0A0L8RKW5 A0A0L8RKW5_SACEU	ECCKEGLDFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1236
tr A0A0L8VRU3 A0A0L8VRU3_S9ACH	ECCKEGLDFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1235
tr J8PP58 J8PP58_SACAR	ECCKEGLDFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1235
tr A0A0C7MY71 A0A0C7MY71_S9ACH	KCKEGLDFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1181
tr C5DNP3 C5DNP3_LACTC	KCKEGLDFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1176
tr C5DX79 C5DX79_ZYGRC	ECGRNGLDFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1197
tr G8ZR00 G8ZR00_TORDC	ACGEEGLDFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1207
tr A0A1E4RQF7 A0A1E4RQF7_SASCO	ACGDDNLFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1203
tr A0A0L0P4K6 A0A0L0P4K6_SASCO	ACGDDNLFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1192
tr A0A1A0HGT7 A0A1A0HGT7_SASCO	ACGREELFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1195
tr C4YB23 C4YB23_CLAL4	ACGDDNLFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1198
tr G3B4C1 G3B4C1_CANTC	TCYKVENISFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1169
tr A3LX46 A3LX46_FICST	ACGDAFLFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1141
tr A0A1E4SMT6 A0A1E4SMT6_SASCO	SCGKGLSFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1143
tr A5DM52 A5DM52_FICCG	ACGQGLSFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1190
tr B5RTF6 B5RTF6_DBBHA	ACCAESGLSFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1199
tr G3AEY0 G3AEY0_SPAFN	ACGERGLKFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1176
tr G8B7X1 G8B7X1_CANPC	SCGESDLQFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1180
tr H8X1L6 H8X1L6_CAN09	SCGESDLQFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1179
tr B9W6L5 B9W6L5_CANDC	KCTENDLQFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1188
tr C4YFJ1 C4YFJ1_CANAW	KCAEFDLQFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1168
tr C5ME71 C5ME71_CANTT	QCCEHLNFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1181
tr M3IK19 M3IK19_CANMX	KCGEALNFAVVRVVRVVAADVDVDMNIIIRLDCP IELH	QVLDVLRKEIEFERRYKQFLV	1186

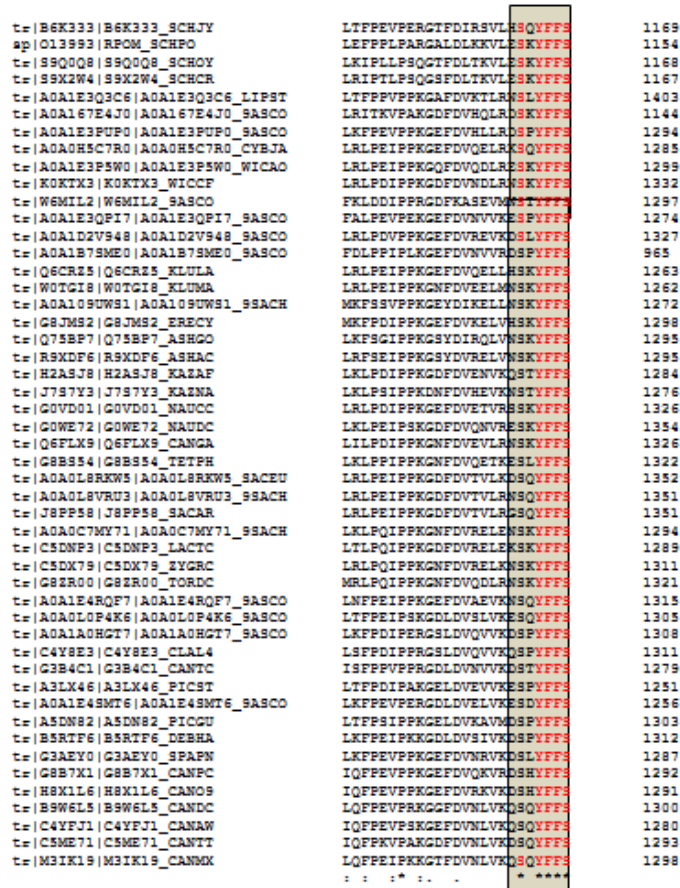


Fig. 5. MSA of various SSU RNA polymerases from various fungal mitochondria

*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 sincaudum*
- tr|G8JMS2|G8JMS2_EREY, *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, *Naumovozya 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, *Torulaspota 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 Mg²⁺ 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 ARATH	-----MPLLLF-----PISPPCVPP	15
AAD03373.1 OSATIWA	-----MSAAAAASPISINPTSHFQ	20
sp P92969 RPOT1_ARATH	MWRNILGRASLRKVKFLSDS-----SSSGTHYPVNRVR-----G-----ILS	37
sp Q93Y94 RPOT1_NICSY	MWRVYISKQAYSRRKFRNSHDSALLGFSQYSSSSFGKTRPLQCLCEESTINPNLGLSQNSIFS	60
sp P06221 RPOL_BESP6	-----	0
sp P18147 RPOL_BFK11	-----	0
ACY75835.1 T7	-----	0
CAC86264.1 T3	-----	0
BAC98394.1	FRPR-LRRLSPPPFMAAVAPP---SL-----STPVTILPSVSVVALLPFLFPATD	60
AAD03373.1	HQTSLVTWIKPPPSALFRKRTLFFFERHSLPISSASSSSSSSTSLSVHEKPISSNV-	78
sp P92969 RPOT1_ARATH	-----SVNLSGVRNGLS-IPVNVEMGGL-----	59
sp Q93Y94 RPOT1_NICSY	R-----ISRKRVRHLEGICESS-RNPHLGLSQNSLFSVVG	95
sp P06221 RPOL_BESP6	-----	0
sp P18147 RPOL_BFK11	-----	0
ACY75835.1 T7	-----	0
CAC86264.1 T3	-----	0
BAC98394.1	DFHWL-----DLFAFLNSPADSYQIPVEEQEVEVEVEV---EVGVERERERE-----	104
AAD03373.1	-----HFHG--NLIESFENQDSSY-----AGTIKASLIEELENFVERNGLSGR	120
sp P92969 RPOT1_ARATH	-----SSFRHQCYVVEGYATAAQAIDSTDPEDES SSGSDEVNELITEMEETER--I--R	110
sp Q93Y94 RPOT1_NICSY	DFRVCGKRGSGSLGFLRSYGSAAEAIASTSEE---DIDEIQELIEEMKNEA--L--K	147
sp P06221 RPOL_BESP6	-----	0
sp P18147 RPOL_BFK11	-----MNAL--NIGRNDF--	11
ACY75835.1	-----MNTI--NIAKNDF--	11
CAC86264.1	-----MNIIE--NIEKNDF--	12
BAC98394.1	-----RERERERARAKAEHRRLR	138
AAD03373.1	RRLFMQDPFWS--ALFLKGLSKMWDQTLKIERKDIDKRFDSLRR	178
sp P92969 RPOT1_ARATH	KKARIA-----AIPFKRVIAGMGAQKFFYMLK	151
sp Q93Y94 RPOT1_NICSY	TNLQ-----FQPKTIIGMVGKYNLLR	185
sp P06221 RPOL_BESP6	-----MQDL-HAIQLQLEEMFMNGIIRRFADQ	41
sp P18147 RPOL_BFK11	SEIELAAIPYNTLSEHYGDQA-AREQLALEHEAYELGRQRFRLMLE	69
ACY75835.1	SDIELAAIPFNTLADHYGERL-AREQLALEHESYEMGEARFRMFL	69
CAC86264.1	SEIELAAIPFNTLADHYGSAL-AREQLALEHESYELGERRFLRML	70
BAC98394.1	DEYRELEREMLDRLAPALPVVKSFLGWFEFLRDAIARDQEVQR--RKRVKHVYAKYLL	196
AAD03373.1	DEYRDLKEMCEKRNLPNLPVVRHMLFGWFLKDVIEREQKLRNKSKKVRAAYAPHIE	238
sp P92969 RPOT1_ARATH	RECREILADMCEQKLPAPNLPYMKSLFLGWFEFVRMAIQDLDLTFK--IKKGRIPYAPFME	209
sp Q93Y94 RPOT1_NICSY	KEYQELMMDCEQKLPAPNLPYMKSLFLGWFEFLRDAIAAEQKLCD--EGKNGRAYAPFQC	243
sp P06221 RPOL_BESP6	-----NRRLSELIAARAEQIQAQYKEEYEGKKGAPRALAFLQC	80
sp P18147 RPOL_BFK11	-----AKELVLTLLPQLTRRIDWKEEQANARGKFRAYYPIKH	108
ACY75835.1	-----AKELITLLPFIARINWFEEVKAARGKRFRTAFQFLQE	108
CAC86264.1	-----AKFLLATLLPKLTTRIVEWLEEYASKKGRKPSAYAPLQL	109
BAC98394.1	IL-----PADKVAIVMHFMMGLMSKDGVASVRVQAAHCIGEA	237
AAD03373.1	LL-----PADKMAIVMHFMMGLVMSGH-EDGCIQVVQAAVSGIA	278
sp P92969 RPOT1_ARATH	QL-----PADKMAVITMHFMMGLMNTNAGSVGIVKLVNAATQIGEA	250
sp Q93Y94 RPOT1_NICSY	QL-----PADKMAVITMHKLMGLMTG-GGTGSARVVQAAASHIGEA	283
sp P06221 RPOL_BESP6	V-----ENEVAAYITMKVVMMLNT--DA---TLQAIAMVAER	114
sp P18147 RPOL_BFK11	GVA3ELAVSMGAEVLKKEKRGVSSAETALLTIEVVLGNHRPLRG---HNEAVSSQLGKA	164
ACY75835.1	I-----KPEAVAYITIKTLCLTSDANT---TVQAVASAIIGA	144
CAC86264.1	I-----KPEASAFITLKVILASLTSTNMT---TIQAAAGMLGKA	145
BAC98394.1	VEREFKQVTFPQTRKK3AGENDL-----ALEKEQAKCRKRVKSLVRRRRLTEA-	286
AAD03373.1	IEQEVRIHNFLEKTRKNNAGDSQE-----ELKERQLLRKRVNSLIRRKRIIDA-	326
sp P92969 RPOT1_ARATH	VEQEVRIINSFLQKKNKNAIDDKTINTEAENVSEEIVAKETEKARKQVTVLMKKNLQV-	309
sp Q93Y94 RPOT1_NICSY	IEHEARLHRFLEKTKKSNALSGLDLETPG-----DIMKERERVRKRVKILMKKQLQV-	337
sp P06221 RPOL_BESP6	IEDQVRF3KLEGHAAKYFEK-----VKS3LKA3R-TK3YRHAHVAVV	156
sp P18147 RPOL_BFK11	IEDEARFGRIRREQAAYFKK-----NVADQLDKRVGHVYKKA-FMQVV	206
ACY75835.1	IEDEARFGRIRDLEAKHFKK-----NVEEQINRKHVGHVYKKA-FMQVV	186
CAC86264.1	IEDEARFGRIRDLEAKHFKK-----HVEEQINRKHGQVYKKA-FMQVV	187

BAC98394.1	----QKIVQQEIELEEWGTEISQVKLGTRLIELELLDLSAFVQSEADQTPESSPDIRPAFRH	341
AAD03373.1	----LKV-VKSEGTKFWGRATQAKLGSRLLELLIEAAYVQPFLLTQSGDSIPEFRPAFRH	380
sp P92969 RPOT1_ARATH	----KALVRRKHSFKFWGQEAQVQVGRALIQLIMENAYIQFPAEQDQDDGPPDIRPAFRQ	364
sp Q93Y94 RPOT1_NICSY	----RKIVRQQDDEKFWQDMLVQVGRCLIQILMETAYIQPRNDQLDDCFFDIRPAFVH	392
sp P06221 RPOL_BPSP6	AEKSVAEKADDFDRWEAWFKETQLQIGTTLLEILEGSAVYFNGEVPVFR	204
sp P18147 RPOL_BPK11	EADMISKMGMLGGINWASWKTDEQMHVGTLLLELLIEGTGL--VEMTK-----	251
ACY75835.1	EADMLSKGLLGGEAWSSWHKEDSIHVGVRCIEMLIESTGM--VSLHR-----	231
CAC86264.1	EADMIGRGLLGGEAWSSWDKETTMHVGIIRLIEMLIESTGL--VELQR-----	232
	* : : * : : *	
BAC98394.1	VLRQPIV-ENGRLEKHKHVVIECDPLVHEGFESTA--RHVEIPYLPMLVTPKKRWGYDTG	397
AAD03373.1	RFKTVIKYPGSKLVRRYGVIECDSELLLAGLDKSA--KHMLIPYVPMPLVTPKKRWGYDTG	437
sp P92969 RPOT1_ARATH	NFRIVTL-ENTKTSRRYGCIECDPLVLEGLDKSA--RHMVIPYLPMLIPQNWITGYDQG	420
sp Q93Y94 RPOT1_NICSY	TLKTV--ETMKGSRRYGVICQDPLVREGLDKTA--RHMVIPYVPMPLVTPQSWLGYDQG	446
sp P06221 RPOL_BPSP6	NMRT----YGGKTIYYL--QTSESVGQWISAFKEHVAQLSPAYAPCVIPPRFWKTPFNG	257
sp P18147 RPOL_BPK11	NKMA----DGSDDVTSMQMVQLAPAFVELLSKRAGALAGISPMHQPCVVPFKPWVETVGG	307
ACY75835.1	QWAG----VVGQD--SETIELAFYAEAIATRAGALAGISPMFQPCVVPFKPWVETVGG	284
CAC86264.1	HNAG----NAGSD--HEALQLAQEYVDVLAKRAGALAGISPMFQPCVVPFKPWVETVGG	285
	: : : : : * : : * : : *	
BAC98394.1	GYLFL--PSYIMRTHGVDKQEAIKSVPRKQLRKVFEALDTLIGSTKWRVNRVHNAVET	454
AAD03373.1	GYLFL--PSYIMRTHGSKKQDALKDISHKTAKHRVFEALDTLIGTKWRVNRNILDVVER	494
sp P92969 RPOT1_ARATH	AHFFL--PSYVMRTHGAKQQRTVMKRTPKQLEPVYEALDTLIGTKWKINKKVLSDVDR	477
sp Q93Y94 RPOT1_NICSY	AYLFL--PSYIMRTHGAKQREAVKRVPRKQLEPVFQALDTLIGTKWRVLRKRVLSIVDR	503
sp P06221 RPOL_BPSP6	GFHTKVASRIRLVKQ--NREHVRLKTKQKMPKVKYKAINALQNTQWQINKDVLAVIEE	313
sp P18147 RPOL_BPK11	GYWSVGRRELALVRTH--SKKALRRYADVHMPEVYKAVNLAQMTQWVNRKVLAVVNE	363
ACY75835.1	GYWANGRRPLALVRTH--SKKALMRYEDVMPEVYKAINIAQNTAWKINKKVLAVANV	340
CAC86264.1	GYWANGRRPLALVRTH--SKKGLMRYEDVMPEVYKAVNLAQNTAWKINKKVLAVVNE	341
	.. : : : : * : : * : : *	
BAC98394.1	I--WSRGGGIA--GLVDKENIPLPERPET-----EDPDEIQKRWKSLKK	494
AAD03373.1	L--WADGGNIA--GLVNRDVP IPEKPS-----EDPEELQSWKWSARK	534
sp P92969 RPOT1_ARATH	I--WANGGRIG--GLVDREDVPIPEEER-----EDQEKFKNWRWESKK	517
sp Q93Y94 RPOT1_NICSY	I--WASGGRLA--DLVDREDVPLPEEEDA-----EDEAQIRKRWKVKVG	543
sp P06221 RPOL_BPSP6	VIRLDLGYGVPSFKPLIDKENKPNPVPVVEFQHLRGLREKEMLSPEQWQFINWEGECAR	373
sp P18147 RPOL_BPK11	IVNWRKCP--VGDVPAIEREELPRFPDDIDTN-----EVARKAWRKEAAA	406
ACY75835.1	ITKWKHCP--VEDI PAIEREELPMKPEDIDMN-----PEALTAWKRAAAA	383
CAC86264.1	IVNWRKCP--VADIPSLERQELPRFPDDIDTN-----EAAIKRWKKAAG	384
	: : : : : * : : *	
BAC98394.1	AKKANRELHAERCDTELKLSVARRMREEDGFYYPHNIDFKGRAYPMHAHLSHLGSDLCRG	554
AAD03373.1	ANKINRERHSLRCDVELKLSVARRMKDEEGFYYPHNIDFKGRAYPMHFLNHLGSDLCRG	594
sp P92969 RPOT1_ARATH	AIKQNRERHSQRCDIELKLEVARRMKDEEGFYYPHNIDFKGRAYPIHFIYLNHLGSDLCRG	577
sp Q93Y94 RPOT1_NICSY	VKKNCRERHSQRCDIELKLEVARRMKDEEGFYYPHNIDFKGRAYPMHFIYLNHLGSDLCRG	603
sp P06221 RPOL_BPSP6	LYTAETKRGSKSAAVVRMVGQARKYSAFESIYFYVAMDSLRVYVQSTLSPQSNLDLGA	433
sp P18147 RPOL_BPK11	VYRKDKARQSRRCCEFMVAQANKFANHKAIWFPYNDWRGRVYAV--SMFNPQGNMTRG	465
ACY75835.1 T7	VYRKDKARQSRRISELFMLEQANKFANHKAIWFPYNDWRGRVYAV--SMFNPQGNMTRG	442
CAC86264.1 T3	IYRLDKARVSRRISELFMLEQANKFASFKAIWFPYNDWRGRVYAV--PMFNPQGNMTRG	443
	: : : : * : : * : : *	
BAC98394.1	VLEYAEGRFL--GKSGLRWLKIHLANKYGGGIEKLSHEDKVAFVEN--QLPDI FDSATNFV	611
AAD03373.1	TLEFAEGRFL--GKSGLHWLKIHLANLYAGGVKLSHDARLAFVEN--HLDDIMD SAENFI	651
sp P92969 RPOT1_ARATH	ILEFCEGRFL--GKSGLRWLKIHIANLYAGGVKLAYEDRIAFTES--HLEDIFDSSDRFL	634
sp Q93Y94 RPOT1_NICSY	ILEFAEGRFL--GKSGLRWLKIHLANVYGGGVKLSYEGRVAFSEN--HVEDIFD SAERFL	660
sp P06221 RPOL_BPSP6	LLRFTDGRFL--NGVEALKWFCINGANLW--GWDKKTFDVRVSNVLDDEFPQDMCRDIAADFL	491
sp P18147 RPOL_BPK11	SLTLARGRFL--GLDGFYWLKIHGANCA--GVDKVPPFERIKFIEE--NEGMI LASAADFL	520
ACY75835.1	LLTLARGRFL--GREGYWLKIHGANCA--GVDKVPPFERIKFIEE--NHENIMACAKSFL	497
CAC86264.1	LLTLARGRFL--GEEGFYWLKIHGANCA--GVDKVPPFERIAFIEK--HVDDILACAKDPI	498
	* : : * : : * : : * : : *	
BAC98394.1	DGNCWMMNAEDPFPQCLAACMDLSDALKSS---SPQCAVSHLPIDGSGCNGLQHYAALGR	668
AAD03373.1	HGKRWWLKAEDPFPQCLAACVILTQALKSP---SPYSVISHLPIDGSGCNGLQHYAALGR	708
sp P92969 RPOT1_ARATH	EGKRWWLNAEDPFPQCLAACINLSEALRSP---FPEAAISHIPIDGSGCNGLQHYAALGR	691
sp Q93Y94 RPOT1_NICSY	EGKRWWLGAEDPFPQCLATCINIAELRSP---SPETAISYMPIDGSGCNGLQHYAALGR	717
sp P06221 RPOL_BPSP6	T--FTQWAKADAPYEFILAWCFEYAGVLDLVDDEGRADEFRTHLPVHDGSGCSGIQHSAMLR	550
sp P18147 RPOL_BPK11	N--NTWWTQDQSPFCFLAFCEYAGVVKH----HGLNMYNCSLPLAFDGS CSGIQHSAMLR	574
ACY75835.1 T7	E--NTWAAEQDQSPFCFLAFCEYAGVQH----HGLSYNCSLPLAFDGS CSGIQHSAMLR	551
CAC86264.1 T3	N--NTWAAEQDQSPFCFLAFCEYAGVTH----HGLSYNCSLPLAFDGS CSGIQHSAMLR	552
	* : : * : : * : : * : : *	

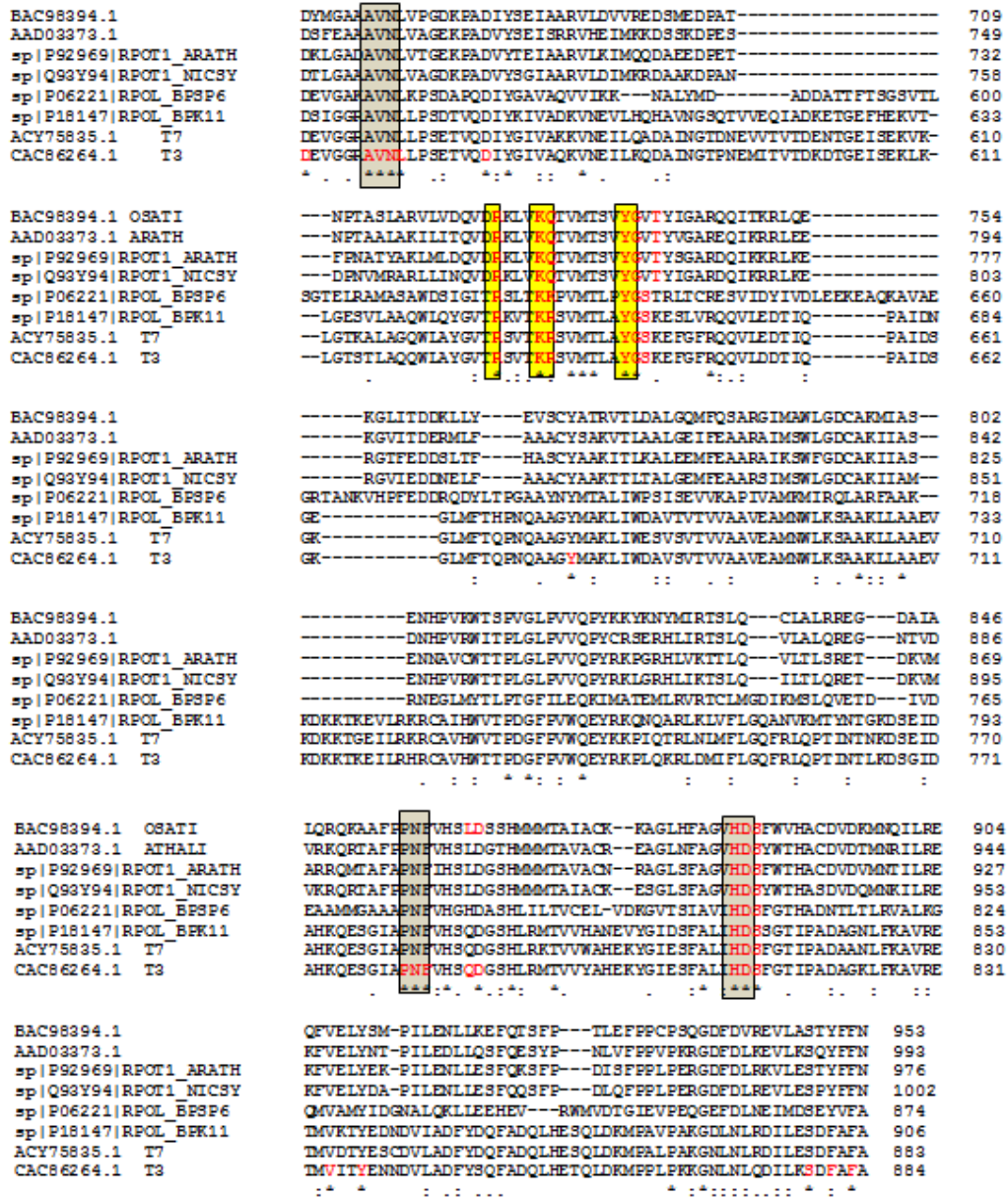


Fig. 6. MSA of T3, T7, K11, SP6 with SSU mitochondrial and chloroplast RNA polymerases

BAC98394.1, *Oryza sativa*, Japonica group (Chloroplast)
AAD03373.1, *Arabidopsis thaliana* (Chloroplast)
sp|P92969|RPOT1_ARATH, *Arabidopsis thaliana* (Mitochondria)
sp|Q93Y94|RPOT1_NICSY, *Nicotiana glauca* (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 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

sp P00582 DPO1_ECOLI	-----MVQIQPNPLILVDGSS-----YLY--RAYHAFPPPLTM	30
sp P06221 RPOL_BPS P6	-----MQDLHA IQLQLEE EMFNGGI RRF EADQQRQI	31
sp P18147 RPOL_BPK11	-MNALN IGRNDFS EIELAAI PFNITLSEHYGDQAAR EQLALEHEAYELGR QRFLMGLERQV	59
ACY75835.1 T7	-MNTIN IAKNDFS DIELAAI PFNITLADHYGERLAR EQLALEHESYEMGE ARFRKMFERQL	59
CAC86264.1 T3	MNI IEN IEKNDFS EIELAAI PFNITLADHYGSALAK EQLALEHESYELGER RRF LKMLE RQA	60
:		
:		
sp P00582 DPO1_ECOLI	SAGEPTGAMYGVNLNMLRSLIMQYKPTHA AVVFDAGKGTFRDE LFEHYKSHRPFMPDRLA	90
sp P06221 RPOL_BPS P6	AAGSES DTAW---NRRL SELIAPM-----AEGTQA YKEEYEGKGRAPRALA-	76
sp P18147 RPOL_BPK11	KAGEFADNAA---AKPLV LTLHPQL-----TKRIDD WKEEQANARGKFPRA YY-	104
ACY75835.1 T7	KAGEVADNAA---AKPLI TTLHPFM-----IARIND WFEEVGA KRGRPTA FQ-	104
CAC86264.1 T3	KAGEIADNAA---AKPLL ATLLPKL-----TTRIVE WLEEYASKGRKPSA YA-	105
** . . : : * :		
* . : : *		
sp P00582 DPO1_ECOLI	QIEPLHAMVKGMLP LLA VSGVEADIVI GTLAREAKGGRPV LITSGDKIDMAQIVTPNIT	150
sp P06221 RPOL_BPS P6	---FLQCV-----	81
sp P18147 RPOL_BPK11	---PIKHG-----	109
ACY75835.1 T7	---FLQEI-----	109
CAC86264.1 T3	---PIQLL-----	110
:		
:		
sp P00582 DPO1_ECOLI	LIDNMTNTILGPE EVVWNYGVVPEL IID--FLALMGDS----SDNIPGVPG-VGERT-AQ	202
sp P06221 RPOL_BPS P6	-----ENEVAAY ITMKVWIMIMLNT--DATLQAI AMSVAERIEDQ	118
sp P18147 RPOL_BPK11	-VASELAVMSGAEVLKPKRQVSEALALITIKRVLGNHRPL KGHNPVAVS QLKGAL EDE	168
ACY75835.1 T7	-----KPEAVAY ITIKTTLACLTSADMTT VQAVASAI GRAI EDE	148
CAC86264.1 T3	-----KPEASAF ITILKVLASLTSNMTT IQAAGMLGCAI EDE	149
* : : : . . . :		
:		
sp P00582 DPO1_ECOLI	ALLQQLGGLDITLYAEPE-----KIAGLSFRGAKTMAAKLE---Q-----	238
sp P06221 RPOL_BPS P6	VRFSKL EGHAAKY FEKQVKS LKASR-TK SYRHANNVAVVAEK SVAEKDADFTIRWEAW PKE	177
sp P18147 RPOL_BPK11	ARFGRI HEQEAA YFKGNVAD QLDKRVGHVYKGA-FMQVVEADMISKGLGGDNWASWKT D	227
ACY75835.1 T7	ARFGRI RDLEAKH FKGNVEE QLNKRVGHVYKGA-FMQVVEADM LSKGLLGGEAWS SWHKE	207
CAC86264.1 T3	ARFGRI RDLEAKH FKGNVEE QLNKRVGHVYKGA-FMQVVEADM IGRGLLGGEAWS SWDKE	208
. . : : : : : : : : * : :		
:		
sp P00582 DPO1_ECOLI	-----NKEWAYLS---YQLA TIRITDVE LELTCEQL E VQQAEEEL	275
sp P06221 RPOL_BPS P6	TQLQIGITLLEILEGSVFYNGEPVEMRAMRTYGGKTIYYL---QTSESVGGWISAPKHEV	234
sp P18147 RPOL_BPK11	EQMHVGTKLELL IEGTGL---VEMTKNKGADGSD DVTSMQVQLAPAFVELLSKRA GAL	284
ACY75835.1 T7	DSIHVGVRCIEML IESTGM---VSLHRQVAGVWQD---SET IELAPEYAEAIATRAGAL	261
CAC86264.1 T3	TTMHVG IRLIEML IESTGL---VELQRHNAGNAGS D---HEALQLAQEYVVDVLA KRAGAL	262
: : : : : : : : :		
:		
sp P00582 DPO1_ECOLI	LGLEFK-----YE FKRWTADVEAGW LQANGAKPAAKPQETS-----VADEAPEVTAT VI	325
sp P06221 RPOL_BPS P6	AQLSPA YAPCVIP PRPWRT P FNGGFHTE KVASRIRLVKGNRE HVKRLTQKQMPKVVYKAIN	294
sp P18147 RPOL_BPK11	AGISPMHQCVVP PKPWVET VGGGYWV GRRELALVRTHSEKALRRYADVHMPEVYKAVN	344
ACY75835.1 T7	AGISPMFQCVVP PKPWVGI TGGGYWAN GRRELALVRTHSEKALMRYEDVHMPEVYKAVN	321
CAC86264.1 T3	AGISPMFQCVVP PKPWVAI TGGGYWAN GRRELALVRTHSEKGLMRYEDVHMPEVYKAVN	322
: * : : * : : : * * : :		
:		
sp P00582 DPO1_ECOLI	SYDNYVITLDEET LKAWIAKLEKAVFA FDTETDS LDNISANLVGLSFA IEPGVAAY IPV	385
sp P06221 RPOL_BPS P6	A-----LQNTQWQDINKD---VLA VI-----EEVIRLD LGYGVPSFKPL	329
sp P18147 RPOL_BPK11	L-----AQNTFWKINKK---VLA VV-----NEIVNWKHCP--VGDVPA	377
ACY75835.1 T7	I-----AQNTAWKINKK---VLA VA-----NVITVWKHCP--VED IPA	354
CAC86264.1 T3	L-----AQNTAWKINKK---VLA VV-----NEIVNWKHCP--VAD IPS	355
. * : : * * : :		
:		
sp P00582 DPO1_ECOLI	AHDVLD--APDQI SRE----RA-LELLKPLLE-----D-----EKA	414
sp P06221 RPOL_BPS P6	LDKRNK PANFVVE FQHLRGRRELKEMLS PEQWQQF INWKGECARLYTAE TKRGSKSAAVV	389
sp P18147 RPOL_BPK11	LEREEL PPRDDI IITN-----EVAR KAWRKEAAAVYRKDKARQSRRCRCE	422
ACY75835.1 T7	LEREEL PKPFEDI IIMN-----PEAL TAWKRAAAAVYRKDKARQSRRI SLE	399
CAC86264.1 T3	LERQEL PPKDDI IITN-----EAAL KEWKGAAAGIYRLDKARVSRRI SLE	400
. * : : : :		

(*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.

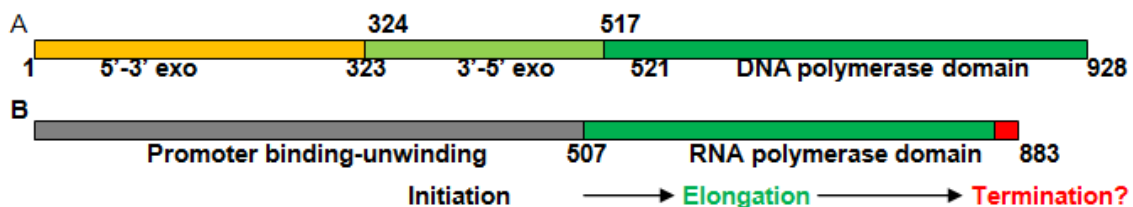


Fig. 8. Dissection of *E. coli* DNA polymerase I (A) and T7 RNA polymerase (B)

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 ATLANTNQLNRK ¹ ILINSLY ⁸ GALGNIH
Human HSV 1	800 AVLLDKQQAIAIK ¹ VVCNSVY ⁸ GFTGVQH
<i>E. coli</i> DNA pol I	748 TVTSEQRRSAK ¹ AINFGLIY ⁸ GMSAFGLAR
<i>E. coli</i> DNA pol II	482 RQGNKPLSQALK ¹ IIMNAFY ⁸ GVLGTTA
<i>E. coli</i> DNA pol III (alpha subunit)	663 YPDVQWQHESLK ¹ PVLEPTY ⁸ GIILYQE
<i>P. furiosus</i> DNA pol	477 KILLDYRQKAIK ¹ LLANSFY ⁸ GYYGYAK
Yeast alpha DNA pol	933 RVQC DIRQQALK ¹ LTANSMY ⁸ GCLGYVN
Human alpha DNA pol	939 ILQYDIRQKALK ¹ LTANSMY ⁸ GCLGFSY
Human Gamma DNA pol	917 TTVGISREHAK ¹ IFNYGRIY ⁸ GAGQPFAER
Human Delta DNA pol (Catalytic subunit)	683 RQVLDGRQLALK ¹ VSANSVY ⁸ GFTGAQV
Human epsilon DNA pol	798 EVLYDSLQLAHK ¹ CILNSFY ⁸ GYVMRKGAR
<i>A. thaliana</i> Delta DNA pol (Catalytic subunit)	679 KAVLDGRQLALK ¹ ISANSVY ⁸ GFTGATV

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

Viral RNA polymerases	
<i>E. coli</i> DNA pol I	-LETVT-----SEQRSAK ¹ AINFGLIY ⁸ GMSAFGLARQLN-----I PRKE-----783
Virus T3	-TSTLAQQWLA ⁸ QVTRSVTK ¹ BSVMTLAY ⁸ GSKEFGFRQQVLD ⁸ DTIQ-----PAIDSGK--664
Virus T7	-TKALAGQWLA ⁸ QVTRSVTK ¹ BSVMTLAY ⁸ GSKEFGFRQQVLE ⁸ DTIQ-----PAIDSGK--663
Virus K11	E SVLAAQWLQ ⁸ QVTRKVT ¹ BSVMTLAY ⁸ GSKESLVRQQVLE ⁸ DTIQ-----PAIDNGE--686
Virus SP6	LRAMASAWDS ⁸ QITRSLTK ¹ KFPVMTLPY ⁸ GSTRLTCRESVIDYIVDLEEKEAQKAVAEGRTA664
Mitochondrial RNA polymerases	
BAC98394.1	AAVNLVPGDKPADYSEIAARVLDVVREDSMEDPATNPTASLARVLVDQVDEKLVK ¹ QIVMTSVY ⁸ GVTYIGARQ 746
AAD03373.1	AAVNLVAGEKPADVYSEISRRVHEIMKKDSSKDPESNPTAALAKILITQVDEKLVK ¹ QIVMTSVY ⁸ GVTYLGARE 786
BAF01496.1	AVNLVAGEKPADVYSEISRRVHEIMKKDSSKDPESNPTAALAKILITQVDEKLVK ¹ QIVMTSVY ⁸ GVTYLGARE 786
BAE98468.1	AVNLVAGEKPADVYSEISRRVHEIMKKDSSKDPESNPTAALAKILITQVDEKLVK ¹ QIVMTSVY ⁸ GVTYLGARE 786

Chloroplast RNA polymerases

t B6K333 B6K333_SCHJY	LLKDRVTRSVKSTVMTNFKVVTYVGARAQIEKQLKLQEDIP-KDLLRDASAFIAKRVFQ	944
ap O13393 RPOCM_SCHPO	FLKDKVTRSVKSTVMTNFKVVTYVGARQISEKLENIIDGME-KLKVADYANYLTKRVFE	937
t S9Q0Q8 S9Q0Q8_SCHRO	VLKDKIDRSVKSTVMTNFKVVTYVGARQIIISQLKRRGDIP-KDMLNYSYLLTKMVFR	951
t S9X2W4 S9X2W4_SCHRCR	ALKDKIDRSVKSTVMTNFKVVTYVGARQIIISQLKRRGDIP-KDMLNYSYLLTKMVFR	950
t A0A1E3Q3C6 A0A1E3Q3C6_LIPST	ILVGVKTRSVKSTVMTNFKVVTYVGARAQIILGQLKDKTKID-ERDLWRCARALTLVLFK	1032
t A0A1E7E4J0 A0A1E7E4J0_SASCO	MAVDKLSRKLKSTVMTNFKVVTYVGARQIISNRLSDA-GLE-QEHLYSTAGYLAKTVLG	801
t A0A1E3PUP0 A0A1E3PUP0_SASCO	LIKDKIDRSVKSTVMTNFKVVTYVGARAQIARQLKDLPHIG-PENIFIVASYLTIINVFA	1007
t A0A0H5C7R0 A0A0H5C7R0_CYBJA	VLKDKIDRSVKSTVMTNFKVVTYIGATAIDKQLADVFPGE-DTY--KYSYLTKRVFA	974
t A0A1E3P5W0 A0A1E3P5W0_WICAO	TLKDKIDRSVKSTVMTNFKVVTYIGATHQIHKQLQDVFPDDT-ESY--KLSYLAKRVFA	982
t K0KTX3 K0KTX3_WICCF	MLKDNIDRSVKSTVMTNFKVVTYVQVATNQHKKLQNVFSED-QSY--KLSYLTKRVFA	1009
t W6MIL2 W6MIL2_SASCO	ILKDLIDRSVKSTVMTNFKVVTYMGASQIARRLEDLGFSPKDDAK--LHRYLARVFD	1015
t A0A1E3QPI7 A0A1E3QPI7_SASCO	KIIPILKRIKSTVMTNFKVVTYIGGAEQIKKQLDAHFDDK-EAY--ALSRFLARVFA	992
t A0A1D2V948 A0A1D2V948_SASCO	LVQHSIKRKLKSTVMTNFKVVTYLGATQIARQLTDEFGKD-TAY--FLSKYLAVRVFA	1035
t A0A1B7SME0 A0A1B7SME0_SASCO	LVKDVLSRKLKSTVMTNFKVVTYVGARAQITRKRKIDIEFDEKYS--MSSKYLTKRVFA	683
t Q6CR25 Q6CR25_KLULA	ILKDLIDRSVKSTVMTNFKVVTYVGAADQIMKELDQVDFDKPEESN--ELSRYLAKRVFA	974
t W0TG18 W0TG18_KLUMA	ILKDLVSRKSTVMTNFKVVTYVGAADQIMKELDQVDFDNPEESN--ELSRYLAKRVFA	973
t A0A1O9UWS1 A0A1O9UWS1_SASCH	LLQDKITRSTVMTNFKVVTYVGAADQIMKELDQVDFDPADDCY--ALSRYLAKRVFA	984
t G8JMS2 G8JMS2_ERECY	QLKDMIDRSVKSTVMTNFKVVTYVGAADQIMKELDQVDFDPADDCY--DMARYLTKRVFA	1014
t Q75BP7 Q75BP7_ASHGO	QLKDLIDRSVKSTVMTNFKVVTYVGAADQIMKELDQVDFDPADDCY--DMARYLTKRVFA	1009
t R9XDF6 R9XDF6_ASHAC	QLKDLIDRSVKSTVMTNFKVVTYVGAADQIMKELDQVDFDPADDCY--DMARYLTKRVFA	1009
t H2ASJ8 H2ASJ8_KAZAF	FLVDKIDRSVKSTVMTNFKVVTLLGATLQIDKQLNDLFDSDSDSM--KYSYLAKRVFA	995
t J7S7Y3 J7S7Y3_KAZANA	FLVDKIDRSVKSTVMTNFKVVTLLGATLQIDKQLNDLFDSDSDSM--KYSYLAKRVFA	986
t G0VD01 G0VD01_NAUDC	ILVDKIDRSVKSTVMTNFKVVTYVGAATQIEKQLSFKFFPDRKCF--DLSKYLTKRVFA	1038
t G0WE72 G0WE72_NAUDC	ILVDKIDRSVKSTVMTNFKVVTYVGAATQIEKQLSXYFFPDRKCF--DLSKYLTKRVFA	1066
t Q6FLX9 Q6FLX9_CAMGA	ILKGVKTRSVKSTVMTNFKVVTYVGAATQIEKQLTAIFDDRAYSL--ELSKYLAKRVFA	1036
t G8B9S4 G8B9S4_TEPFH	VLQDKITRSTVMTNFKVVTYVGAATQIEKQLSNVDFDNPEESN--ELSKYLTKRVFA	1034
t A0A0L8RKS5 A0A0L8RKS5_SACEU	ILKDKIDRSVKSTVMTNFKVVTYVGAATQIEKQLSFIFFDRKESL--DPSKYLTKRVFA	1061
t A0A0L8VRU3 A0A0L8VRU3_SASCH	ILKDKIDRSVKSTVMTNFKVVTYVGAATQIEKQLSFIFFDRKESL--DPSKYLTKRVFA	1060
t J8P5S8 J8P5S8_SACAR	LLQDKITRSTVMTNFKVVTYVGAATQIEKQLSFIFFDRKESL--DPSKYLTKRVFA	1060
t A0A0C7MY71 A0A0C7MY71_SASCH	LLKTMIDRSVKSTVMTNFKVVTYVGAATQIEKQLSNVDFDNPEESN--EMSKYLTKRVFA	1006
t C5DNP3 C5DNP3_LACTC	LLKDKIDRSVKSTVMTNFKVVTYVGAATQIEKQLSFIFFDRKESL--ELSKYLTKRVFA	1001
t C5DX79 C5DX79_EYGRG	ILKDKIDRSVKSTVMTNFKVVTYVGAATQIEKQLSNIFFDRKESL--ELSKYLTKRVFA	1022
t G8ER00 G8ER00_TORDC	TLKDKIDRSVKSTVMTNFKVVTYVGAATQIEKQLSHIIFDRRTYSL--ELSKYLTKRVFA	1032
t A0A1E4RQF7 A0A1E4RQF7_SASCO	FFDKIDRSVKSTVMTNFKVVTYVGAATQIEKQLVHHPKDEDEKARKFTKYLTSLVFD	1028
t A0A0L0F4K6 A0A0L0F4K6_SASCO	FLRQKIDRSVKSTVMTNFKVVTYVGAATQIEKQLSALFKGEPSEKLSIYSRYLTKRVFA	1017
t A0A1A0HG77 A0A1A0HG77_SASCO	FLRQKIDRSVKSTVMTNFKVVTYVGAATQIEKQLSGLFKGDDYATVQKHSRYLTSLVFA	1020
t C4Y8E3 C4Y8E3_CLAL4	FLRQKIDRSVKSTVMTNFKVVTYVGAATQIEKQLSGLFKGDDYATVQKHSRYLTSLVFA	1023
t G3B4C1 G3B4C1_CAMTC	FLRQKIDRSVKSTVMTNFKVVTYVGAATQIEKQLKLVHFDSDSEENSGEYQYLTQHVFA	995
t A3LX46 A3LX46_PICST	FLRQKIDRSVKSTVMTNFKVVTYVGAATQIEKQLDQVFPGE-DTY--EEDVADYARLTKRVFA	967
t A0A1E4SMT6 A0A1E4SMT6_SASCO	FLRQKIDRSVKSTVMTNFKVVTYVGAATQIEKQLDHFHTKDEEDVVDYARLTKRVFA	970
t A5DN82 A5DN82_PICGU	FLRQKIDRSVKSTVMTNFKVVTYVGAATQIEKQLDQVFPGE-DTY--EEDVADYARLTKRVFA	1015
t B5RTF6 B5RTF6_DESHA	FLRQKIDRSVKSTVMTNFKVVTYVGAATQIEKQLDHFHTKDEEDVVDYARLTKRVFA	1024
t G3AEY0 G3AEY0_SAPFN	FLRQKIDRSVKSTVMTNFKVVTYVGAATQIEKQLDHFHTKDEEDVVDYARLTKRVFA	1005
t G8B7X1 G8B7X1_CAMPC	FFDKIDRSVKSTVMTNFKVVTYVGAATQIEKQLVAQHFGEDEHDAH--IYTKYLTKRVFA	1007
t H8X1L6 H8X1L6_CAMO9	FFDKIDRSVKSTVMTNFKVVTYVGAATQIEKQLVAQHFGEDEHDAH--IYTKYLTKRVFA	1006
t B9W6L5 B9W6L5_CAMDC	FLRQKIDRSVKSTVMTNFKVVTYVGAATQIEKQLVAQHFGEDEHDAH--MPARYLARVFA	1016
t C4YFJ1 C4YFJ1_CAMAW	FLRQKIDRSVKSTVMTNFKVVTYVGAATQIEKQLVAQHFGEDEHDAH--AAARYLARVFA	996
t C5ME71 C5ME71_CAMTT	LLQDKITRSTVMTNFKVVTYVGAATQIEKQLSFIFFDRKESL--LYARYLARVFA	1011
t M3IK19 M3IK19_CAMMX	FLRQKIDRSVKSTVMTNFKVVTYVGAATQIEKQLSFIFFDRKESL--LYARYLARVFA	1013

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 polymerase I, DNase footprinting assay using DNase I and methidium-propyl EDTA-Fe²⁺ 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 Tyr⁷⁶⁶ at the active site of the enzyme [11]. Thus, the foot printing and photo cross-linking experiment has suggested the Tyr⁷⁶⁶ 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
<i>E. coli</i> pol I	Arg754/Lys758	Tyr766/Gly 767	Asp705/Asp882
<i>E. coli</i> pol II	Leu*523/Arg527	Phe533/Gly534	Asp452/ Asp545
<i>E. coli</i> 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[#]			
<i>A Thaliana</i>	Arg725/Lys729	Tyr737/Gly738	Asp654/Asp886
<i>O. Sativa</i>	Arg765/Lys769	Tyr777/Gly778	Asp694/Asp926
Mitochondrial RNA polymerases[#]			
<i>A Thaliana</i>	Arg748/Lys752	Tyr760/Gly761	Asp677/Asp909
<i>N. sylvestris</i>	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 Lys⁷⁵⁸ 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 "Kx3NSxYG" 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 "R⁴xxxK¹xxxxxY⁺⁸G". By using a library with totally random nucleotides at five different codons (R659, R660, K663, F667, and G668), Suzuki et

al. [16] confirmed that R⁶⁵⁹ and K⁶⁶³ were immutable in the DNA polymerase from *T. aquaticus* (R⁷⁵⁴ and K⁷⁵⁸ 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 KYI/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 Y⁷⁶⁶ and Y⁷⁵⁸ 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, Doublet 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 Y⁹⁵⁵ residue is critical for nucleotide recognition among Family A DNA polymerases, i.e., γ polymerases from eukaryotes. Furthermore, Y⁹⁵⁵ is a highly conserved residue among a wide variety of DNA polymerases (Table 1). Further proof of Y⁷⁶⁶ involvement in nucleotide selection was obtained from site-directed mutagenesis of Y⁷⁶⁶, substitution of an equivalent amino acid as in Y⁷⁶⁶→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 K⁵²² (\equiv K⁷⁵⁸ 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 thumb 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
$K^{172} \rightarrow L$	No change in activity	[28]
$P^{563} \rightarrow A$	Inactivated	[24]
$Y^{571} \rightarrow S$	Inactivated	ibid
$K^{631} \rightarrow G/L/R$	Partially inactivated	ibid
$T^{636} \rightarrow P$	Inactivated	ibid
$Y^{639} \rightarrow D$	Inactivated	ibid
$F^{646} \rightarrow C$	Inactivated	ibid
$D^{537} \rightarrow N$	Inactivated (Total)	[25]
$D^{812} \rightarrow N$	Inactivated (Total)	ibid
$K^{631} \rightarrow M$	Inactivated (1% activity)	[29]
$Y^{639} \rightarrow F$	Fully active*	ibid
$S^{641} \rightarrow A$	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 Ser⁶⁴¹ in T7 RNA polymerase (Y⁶³⁹GS⁶⁴¹) play any role in the discrimination between NTPs and dNTPs, they made a single amino acid substitution and the S⁶⁴¹ 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 (R⁶²⁷, K⁶³¹ and Y⁶³⁹) are found to be directed towards the substrate binding cleft [30]. Temiakov et al. [31] have also shown that Y⁶³⁹ 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

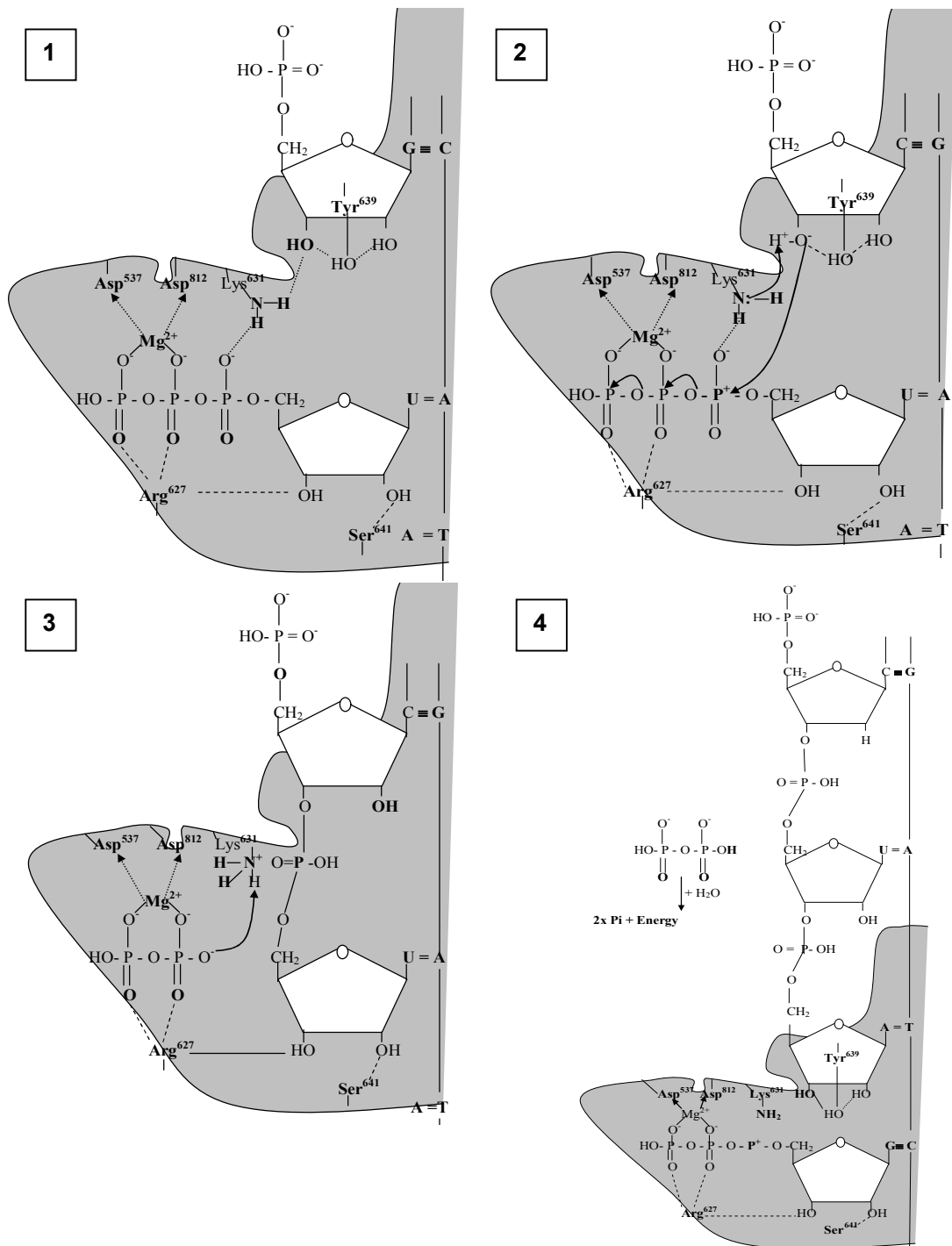
Fig. 9.2. Electronic transition at the active site for proton abstraction and an electrophilic and nucleophilic attack

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 of T7 RNA polymerase that are able to synthesize both RNA and DNA. *FEBS Letters.* 1995;369:165-168.
6. König 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, 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, Senechal 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 pyridoxal 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. Doublet 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'-exonucleolytic 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 mutagenesis 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, Ilgenfritz 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. Doublet 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, Vassilyev 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.

© 2017 P. Palanivelu; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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