

Research paper

DNA polymerases: An insight into their active sites and catalytic mechanism

P. Palanivelu*

Professor & Head, Department of Molecular Microbiology, School of Biotechnology,
Madurai Kamaraj University, Madurai – 625 021, India.

ABSTRACT

Aim: To analyze the active sites of various prokaryotic and eukaryotic polymerases and propose a plausible mechanism of action for polymerases with the *E. coli* DNA polymerase I as a model system.

Study design: Bioinformatics, Biochemical 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 2007 to 2012.

Methodology: The advanced version of T-COFFEE was used to analyze both prokaryotic and eukaryotic polymerase sequences. Along with this bioinformatics data, X-ray crystallographic and biochemical data were used to confirm the possible amino acids in the active sites of different types of polymerases from various sources.

Results: Multiple sequence analyses of various polymerases from different sources show only a few highly conserved motifs among these enzymes except eukaryotic epsilon polymerases where a large number of highly conserved sequences are found. Possible catalytic/active site regions in all these polymerases show a highly conserved catalytic amino acid K/R and the YG/A pair. A distance conservation is also observed between the active sites. Furthermore, two highly conserved Ds and DXD motifs are also observed.

Conclusion: The highly conserved amino acid K/R acts as the proton abstractor in catalysis and the YG/A pair acts as a “steric gate” in selection of only dNTPs for polymerization reactions. The two highly conserved Ds act as the “charge shielder” of dNTPs and orient the alpha phosphate of incoming dNTPs to the 3'-OH end of the growing primer.

Keywords: T-COFFEE; conserved motifs, active sites, eukaryotic polymerases; *E. coli* DNA polymerase I, exonuclease active site, polymerase active site, proof-reading mechanism, polymerization mechanism.

*Corresponding author E-mails: PP@mrna.tn.nic.in/ ppmkupp@gmail.com

1. INTRODUCTION

DNA polymerases are cardinal enzymes, which play a vital role in not only preserving but also maintaining the blue print of life in all living cells. Polymerization of deoxynucleoside triphosphates (dNTPs) into DNA by *E. coli* extract was first demonstrated by Arthur Kornberg in 1955 [1]. Since then a large number of DNA polymerases from various sources have been purified, characterized, cloned and sequenced. Complete nucleic acid and protein sequence data are available for many of these enzymes. These data have become valuable tools in analyzing and understanding the highly conserved motifs and their structure-function relations of these enzymes. Many of these prokaryotic enzymes are usually multifunctional enzymes and exhibit three different activities, viz., polymerization, proof-reading and DNA repair. Some of the polymerases like *Taq* and *Pfu* are extensively used in polymerase chain reaction techniques because of their exceptional thermostability. Except viruses, there is always more than one type of polymerases in a living cell. For example, at least five different polymerases have been characterized in prokaryotes such as *E. coli*, (e.g.), pol I, II, III, IV and V. Eukaryotic DNA polymerase family is a more complex one. At least 18 different polymerases have been reported from eukaryotes. To distinguish them from prokaryotic polymerases, they are called alpha, beta, gamma, delta, epsilon, zeta, eta, theta, iota, tau, etc. Among them, the five key enzymes are α , β , γ , δ and ϵ which are also analyzed in this article. Except γ polymerases, all are located in the nucleus. The DNA polymerases, which undertake replication, are also known as replicases.

1.1 Dynamics of DNA polymerization and Proof reading activities

DNA polymerases synthesize a new strand of DNA on a given template DNA using a primer and the four dNTPs. They add nucleotides one at a time to the 3' OH end of the primer as shown in Fig. 1 and thus the new strand grows from 5' → 3'. The overall reaction catalyzed by DNA polymerases may be written as,

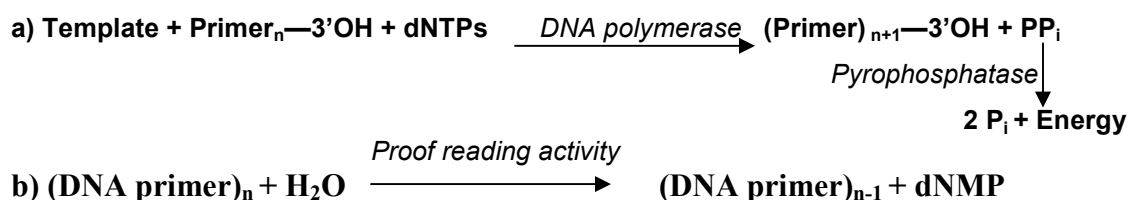


Fig. 1 Dynamics of a) DNA polymerization and b) Proof reading activity

In a dNTP, 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 release of large negative free energy, which is utilized in the formation of the phosphodiester bond involving large positive free energy. The inorganic pyrophosphate formed during the reaction is hydrolyzed immediately by the enzyme pyrophosphatase and results in release of more energy which helps in the translocation of the polymerase on the template. Such a coupling of reaction drives the polymerization reaction only in the forward direction, (i.e.), the polymerized nucleotides cannot be depolymerized.

1.2 Crystal structure of polymerases

Crystal structures of many polymerases like Klenow polymerase, DNA polymerase I of *E. coli*, *Taq* polymerase, HIV-1 Reverse transcriptase, T₄ RNA polymerase, rat DNA polymerase β , etc., are known. Although their primary structures are markedly different (probably due to convergent evolution), their crystal structures are remarkably similar in overall shape. All DNA polymerases, irrespective of their source, have a common protein fold that resembles the shape of a half opened "right hand" with three distinct domains, "thumb", "palm" and "fingers" with analogous functions. The catalytic centre is located on the "palm" domain with two conserved Asp

residues. The “thumb” and “fingers” domains, although analogous, are not homologous between the pol families. The “thumb,” “fingers,” and “palm” form a pocket along which the DNA template with the primer can move. The DNA molecule interacts with specific amino acids located in the “palm” region whereas the “thumb” and “finger” are involved in the polymerization reaction. The finger and thumb domains wrap up the DNA around and hold it across the enzyme’s active site, which comprises residues primarily from the palm domain. Among them the palm region is the most conserved [2]. Similar spatial arrangements of the highly conserved carboxylic acids in these enzymes suggest that the polymerases contain similar active sites and therefore, are likely to operate by the same catalytic mechanism.

2. MATERIALS AND METHODS

For multiple sequence analysis of various polymerases the sequences were retrieved from SWISS-PROT and PUBMED sites and analyzed using the advanced version of T-COFFEE available at the website: <https://tcoffee.org>.

3. RESULTS AND DISCUSSION

3.1 Active site analysis of DNA polymerases

To-date the most well characterized polymerase is the DNA polymerase I of *E. coli*. DNA polymerase I is a multifunctional enzyme and harbours at least 6 different binding/active sites on the single polypeptide viz., One binding site for all four dNTPs, One binding site for template DNA or at nicks on DNAs, One binding site for the primer, One binding site for the 3' OH on the primer, One binding site (exonuclease-1 active site) for proof-reading function and One binding site (exonuclease-2 active site) for DNA repair-function. In addition to the above binding sites, it also contains several binding sites for metal ions as well. The dNMP site (part of proof-reading active site) and dNTP site (part of polymerase active site) bind their respective ligands non-competitively. In eukaryotic polymerases the polymerization and proof reading activities are accomplished by two different polymerases. In order to analyze various active and binding sites of different polymerases a multiple sequence alignment was done.

3.2 Multiple sequence analysis of different polymerases from different sources

Figures 2A – 2E show the results of multiple sequence alignment of various prokaryotic polymerases (pol I to pol V). (The conserved regions are shaded and the active site regions are underlined and shaded yellow).

Figure 2F shows the results of multiple sequence alignment of various viral polymerases. (The conserved regions are shaded and the active site regions are underlined and shaded yellow).

Figures 2G – 2K show the results of multiple sequence alignment of various eukaryotic polymerases (alpha, beta, gamma, delta and epsilon). (The conserved regions are shaded and the active site regions are underlined and shaded yellow).

It is clear from the figures 2A-2K, irrespective of the type and origin of the polymerase, marked similarities are observed in their catalytic regions which are summarized in Table 1.

Fig. 2A Prokaryotic DNA Polymerases (pol I)

sp P52026 DPO1_BACST	GKTTFRHETFDYQKGGROQTPELSEQFPLRELLKAYRIPAYELDHYEADDIIGTMAAR	120
sp P00582 DPO1_ECOLI	KGKTFRDELFEHYKSHRPPMPDDLRAQIEPLHAMVKAMGLPLLAVSGVEADDVIGTLARE	124
sp P0A550 DPO1_MYCTU	SRQTFRLQRYPEYKANRSSTPDEFAGQIDITKEVLGALGITVLSEPGFEADDLIATLATQ	134
sp O34996 DPO1_BACSU	GKTTFRHGTFKEYKGGROKTPELSEQMPFIRELLDAYQISRYELEQYEADDIIGTLAKS	122
sp Q9F173 DPO1_SALTY	KGKTFRDELFEHYKSHRPPMPDDLRAQIEPLHAMVKAMGLPLLAVSGVEADDVIGTLARE	124
sp P59200 DPO1_STRR6	GKTTFRTEMYADYKGGRAKTPDEFREQFPFIRELLDHMGIRHYELAQYEADDIIGTLDKL	125
sp O32801 DPO1_LACLM	GKTTFRTEMFADYKGGRSKTPDEFREQLPFIKEMIEKLGIRHYELANYEADDIIGTLDKM	126
sp P43741 DPO1_HAEIN	KGKTFRDEMFEQYKSHRPPMPDDLRLKQIQPLHDMIRALGIPLLVVEGIEADDVIGTLALQ	124
sp Q9ZJE9 DPO1_HELPJ	QTKTKRAEKLGEYKQNRKDAPKEMLLQIPIALEWLQKMGFTCEVVGGEADDVIASLATL	130
sp Q9CDS1 DPO1_LACLA	GKTTFRTEMFADYKGGRSKTPDEFREQLPFIKEMIEKLGIRHYELANYEADDIIGTLDKM	126
sp P59199 DPO1_STRPN	GKTTFRTEMYADYKGGRAKTPDEFREQFPFIRELLDHMGIRHYDLAQYEADDIIGTLDKL	125
sp Q9S1G2 DPO1_RHILE	SAKTFRKDLIDAYKANRSAPPEELIPQFGLIREATRAFNLPCTIETEGFEADDIATYARQ	147
sp Q92GB7 DPO1_RICCN	GGKNFRHQIYPDYKANRPPPPEDLIQPLPLVRDVASNLNFPILEKNGYEADDIATYATK	122
sp P0A551 DPO1_MYCBO	SRQTFRLQRYPEYKANRSSTPDEFAGQIDITKEVLGALGITVLSEPGFEADDLIATLATQ	134
sp Q9HT80 DPO1_PSEAE	KGPTFRDELFAEYKANRPSMPDDLRLVQVEPLHASVRALGLPLLCVEGVEADDVIGTLARS	121
sp Q04957 DPO1_BACCA	GKTTFRHEAFQYKGGROQTPELSEQFPLRELLRAYRIPAYELENYEADDIIGTLAAR	120
sp P46835 DPO1_MYCLE	SRKTFRSECYAGYKANRSSIPAEFHQIDITKEVLGALGITVFAEAGFEADDLIATLATQ	135
sp P52027 DPO1_DEIRA	PVKTFRHEQYEGYKSGRAQTPEDLPGQINRIRALVDALGFPRLEEPGYEADDVIASTTRM	163
	. * ** * * :: *. :	
		****:*.:
sp P52026 DPO1_BACST	AER--EGFAVKVISGDRDLTQLASPVQTVETTKKGITDIESYTPETVVEKYGLTPQIVD	178
sp P00582 DPO1_ECOLI	AEK--AGRPVLISTGDKDMAQLVTPNITLINTMTN----TILGPEEVVNKYGVPELIIID	178
sp P0A550 DPO1_MYCTU	AEN--EGYRVLVVTGDRDALQLVSDVTVLYPRKGVSELTRFTPEAVVEKYGLTPRQYPD	192
sp O34996 DPO1_BACSU	AEK--DGFEVKVFSQDKDLTQLATDKTTVAITRKGITDVEFYTPPEHVKEKYGLTPQIIVD	180
sp Q9F173 DPO1_SALTY	AEK--VGRPVLIISTGDKDMAQLVTPNITLINTMTN----TILGPEEVVNKYGVPELIIID	178
sp P59200 DPO1_STRR6	AEQ--DGFDITIVSGDKDLIQLTDEHTVVEISKKGVAEFEAFTPDYLMEEMLTPAQFID	183
sp O32801 DPO1_LACLM	AEAPDVNFVDVTIVTGDKDMIQLVDGNTRVEISKKGVAEFEAFTPDYLLKMGMLTPAQFID	186
sp P43741 DPO1_HAEIN	ASS--LGKKVLISTGDKDMAQLVDDNIMLINTMNN----SLDRKGVIEKYGIPPELIIID	178
sp Q9ZJE9 DPO1_HELPJ	SP----YKTRIIYSKDKDFNQLLSDKIALFDGK-----TEFLAKDCVKEYGILPQFTD	179
sp Q9CDS1 DPO1_LACLA	AEAPNVNFDVTIVTGDKDMIQLVDGNTRVEISKKGVAEFEAFTPDYLLKMGMLTPAQFID	186
sp P59199 DPO1_STRPN	AEQ--DGFDITIVSGDKDLIQLTDEHTVVEISKKGVAEFEAFTPDYLMEEMLTPAQFID	183
sp Q9S1G2 DPO1_RHILE	AEA--TGADVTIVSSDKDLMLQVSPNVHMYDSMKD---KQIGIPDVIEKWGPPEKMIID	201
sp Q92GB7 DPO1_RICCN	TAA--LGAHVLIISTGDKDLQLVSDVTVLYPRKGVSELTRFTPEAVVEKYGLTPRQYPD	192
sp P0A551 DPO1_MYCBO	AEN--EGYRVLVVTGDRDALQLVSDVTVLYPRKGVSELTRFTPEAVVEKYGLTPRQYPD	192
sp Q9HT80 DPO1_PSEAE	SAA--ADRPVVISITGDKDMAQLVDGHTLVNTMTG---SRLDVDGVKEFGVGPPELIIID	175
sp Q04957 DPO1_BACCA	AEQ--EGFEVKVISGDRDLTQLASPHVTVDITKKGITDIEPYTPEAVREKYGLTPQIIVD	178
sp P46835 DPO1_MYCLE	AEN--EGYRVLVVTGDRDALQLVSDVTVLYPRKGVSELTRFTPEAVIEKYGVTPAQYVD	193
sp P52027 DPO1_DEIRA	AEG--KGYEVRIVTSRDRDAYQLLDEHVKVIAN----DFSLIGPAQVEEKYGVTVRQWVD	216
	:	: : *.* ** . :
		: : *
sp P52026 DPO1_BACST	NMEFTGVKVDTKRLEQMGAELTEQLQAVERRIYELAGQEFNINSPKQLGTVLFDKQLQLPV	546
sp P00582 DPO1_ECOLI	RIERNGVKIDPKVLHNSHEELTLRLAELEKKAHEIAGEEFNLSSTKQLQTILFEKQGKIP	598
sp P0A550 DPO1_MYCTU	KMESAGIAVDLPMLTELQSQFGDQIRDAEEAAYGVIGKQINLGSFKQLGVVLFDELGMKP	573
sp O34996 DPO1_BACSU	EMESTGVKVDVDRDKRMGEELGAKLKEYEKEIHEIAGEFPFNINSPKQLGVILFEKIGLPV	550
sp Q9F173 DPO1_SALTY	RVERNPKIDPAVLHKKHSEETLRLAELEKKAHDIAAGEAFNLSSTKQLQTILFEKQGKIP	598
sp P59200 DPO1_STRR6	KMEIAGIVVKKETLLEMQAENELVIEKLTQEIYELAGEEFNINSPKQLGVLLFEKLGKLP	545
sp O32801 DPO1_LACLM	KMEITGISVSQNTLEEIGAENEELASLTREIYDLAGEEFNINSPKQLGVILFEKLGKLPV	548
sp P43741 DPO1_HAEIN	RMERTGVILDSDALFMQSNEIASRLTALEKQAYALAGQPPFNLA	601
sp Q9ZJE9 DPO1_HELPJ	GMEFQGFKIDAPYFKRLEQEFKNELNVLERQILDIGVDFNLSSTKQLGCVLYDKLGLPK	575
sp Q9CDS1 DPO1_LACLA	KMEITGIAVSQNTLEEIGAENEELASLTREIYDLAGEEFNINSPKQLGVILFEKLGKLPV	548
sp P59199 DPO1_STRPN	KMEIAGIMVKKETLLEMQAENELVIEKLTQEIYELAGEEFNINSPKQLGVLLFEKLGKLP	545
sp Q9S1G2 DPO1_RHILE	RMEARGITVDRQILSRLSGELAQGAARLEDEIYVLAGERFNIGSPKQLGDILFGKMGSLG	679
sp Q92GB7 DPO1_RICCN	KMEKVGITVDANYLKLQSAEFGTEILKLEELIFALSGTKFNIGSPKQLGEILFEKMQLPF	545
sp P0A551 DPO1_MYCBO	KMESAGIAVDLPMLTELQSQFGDQIRDAEEAAYGVIGKQINLGSFKQLGVVLFDELGMKP	573
sp Q9HT80 DPO1_PSEAE	RIERNGALVDANLLGIQSRELGEKMVALERQAYDLAGEEFNLSSTKQLGAILYDKLGLPV	582
sp Q04957 DPO1_BACCA	EMEFAGVKVDTKRLEQMGEELAEQLRTVEQRIYELAGQEFNINSPKQLGVILFEKLGKLPV	546
sp P46835 DPO1_MYCLE	DMEKAGIAADLRLLTELQSQFGDQIRDAEEAAYAVIGKQINLGSFKQLGVVLFDELGMKP	577
sp P52027 DPO1_DEIRA	RMEVRGVQVDSDFLQTLISIQAGVRLADLESQIHEYAGEEFHISPKQLGTVLYDKLELAS	627
	: * * . :	: * : : *

214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284

Fig. 2A CLUSTAL format for T-COFFEE Version of Pol I DNA polymerases (Prokaryotes)
The possible catalytic region is underlined and shaded yellow. Only regions showing highly conserved blocks are shown.

Only regions showing highly conserved blocks are shown.

Fig. 2C Prokaryotic DNA polymerases (Poll III)

tr Q8G7V8 Q8G7V8_BIFLO	FCGYFLVVADYIQWAKDHGIMV	GPGRGSAAGAMVAYS	MGITELDPLKHGL	LIFERFNLNPER	418
tr C4ZRS6 C4ZRS6_ECOBW	FPGYFLIVMEFIQWSKDNVFPV	GPGRGSGAGSLVAYALKITDLDPLEFD	LLFERFNLNPER		396
tr Q88MG5 Q88MG5_PSEPK	FPGYFLIVMDFIKWAKNNNDVPV	GPGRGSGAGSLVAYVLKITDLDPLAYD	LLFERFNLNPER		398
tr C1DST1 C1DST1_AZOVD	FPGYFLIVMDFIKWAKNNNDVPV	GPGRGSGAGSLVAYALKITDLDPLAYD	LLFERFNLNPER		399
tr Q1D042 Q1D042_MYXXD	FCGYFLIVQDFINWAKKMNIPV	GPGRGSGAGSLVAYALRITDVPDIPYN	LLFERFNLNPER		398
tr Q72ZD3 Q72ZD3_BACC1	FSDYFLIVWDFMKYAHENHILT	GPGRGSAAGSLVSYVLEITDIDPIEYD	LLFERFNLNPER		378
tr Q97RC7 Q97RC7_STRPN	FDDYFLVVWDLRFQGSNGYYM	GMGRGSAVGSLSVSYALDITGIDPVEKN	LIFERFNLNPER		344
tr Q4QMF6 Q4QMF6_HAEI8	FPGYFLIVMEFIQWSKDNNDIPV	GPGRGSGAGSLVAYALKITDLDPLEFD	LLFERFNLNPER		396
tr Q2FG35 Q2FG35_STAA3	FEDYFLIVSDLIHYAKTNDVMV	GPGRGSSAGSLVSYLLGITTIDPIKFN	LLFERFNLNPER		342
tr A1R5G9 A1R5G9_ARTAT	FPGYFLVVADFINWAKNNNDGIRV	GPGRGSGAGSMVAYAMRITDLDPLQHG	LIFERFNLNPER		419
tr Q6NGE6 Q6NGE6_CORDI	YPSYFLIVAEILKHARSIGIRV	GPGRGSAAGALVAYALTITNIDPMEHG	LLFERFNLNPER		416
tr Q6D8C8 Q6D8C8_ERWCT	FPGYFLIVMEFIQWSKDNNDVPV	GPGRGSGAGSLVAYSALKITDLDPLEFD	LLFERFNLNPER		396
tr Q7NZL1 Q7NZL1_CHRVO	FPGYFLIVADFIQWKGNSGCPV	GPGRGSGAGSLVAYSLSITDLDPLKYAL	LLFERFNLNPER		398
tr Q5NM98 Q5NM98_ZYMMO	FPGYFLIVADFIKWAQSNQIPV	GPGRGSGAGSVVAVALTITNLDPLRL	GLLFFERFNLNPER		347
tr Q6KHF7 Q6KHF7_MYCMO	FADYFLIIYDIVKFAKENKISV	GPGRGSAAGSIVSYLLNITSINPLKYD	LLFERFNLNPER		324
tr C3K6H1 C3K6H1_PSEFS	FPGYFLIVMDFIQWAKSNGVPV	GPGRGSGAGSLVAVYVQKITDLDPLEYD	LLFERFNLNPER		398
tr B8ZR78 B8ZR78_MYCLB	FPAYFLIVADLVNARSIGIRV	GPGRGSAAGSLAAYALGITDIDPIPHG	LLFERFNLNPER		412
tr A1F2E4 A1F2E4_VIBCH	FPGYFLIVMEFIQWSKDNNDIPV	GPGRGSGAGSLVAYALKITDLDPLEYD	LLFERFNLNPER		401
tr C7C104 C7C104_HELPH	FPGYMLIVWDFIRYAKEMGIPV	GPGRGSAAGSLVAFALKITDIDPLKYD	LLFERFNLNPER		420
tr Q2K9R0 Q2K9R0_RHIEC	FPGYFLIVSDFIKWAKQHDIPV	GPGRGSGAGSLVAYALTITDVPDLRF	SLLFFERFNLNPER		423
tr B5Y1I7 B5Y1I7_KLEP3	FPGYFLIVMEFIQWSKDNNDVPV	GPGRGSGAGSLVAYALKITDLDPLEFD	LLFERFNLNPER		396
tr Q2KYR1 Q2KYR1_BORAI	FPGYFLIVQDFINWGNKNDVPV	GPGRGSGAGSLVAYALGITDLDPIRYD	LLFERFNLNPER		401
tr Q5HV63 Q5HV63_CAMJR	FSGYMLIVHDFIKVAKDKGIPV	GPGRGSAAGSLVSYCLRITDLDPIPS	LLFERFNLNPER		421
	: *::: : : . :	* * * * . * : : : : * * : : :	* : * * * * : *		
tr Q8G7V8 Q8G7V8_BIFLO	VSLPDIDVD	EDPEGRARVIEYCGEKYGTDKVAQC	VIYGTIKTKQALKDSARIM-GY-EFS		476
tr C4ZRS6 C4ZRS6_ECOBW	VSMPDFDVDF	CMEKRDQVIEHVADMYGRDAVSQI	ITFGTMAAKAVIRDVGRVL-GH-PYG		454
tr Q88MG5 Q88MG5_PSEPK	VSMPDFDVDF	CMDGRDRVIDYVAEAYGRNAVSQI	ITFGTMAAKAVVRDVARVQ-GK-SYG		456
tr C1DST1 C1DST1_AZOVD	ISMPDFDVDF	CMDGRDRVIDYVAEAYGRNAVSQI	ITFGSMAAKAVVRDVARVQ-GK-SYG		457
tr Q1D042 Q1D042_MYXXD	VSMPDFDIDF	CQDRRDEVIQYVGRKYGEMNVGQI	ITFGSLKAKSVLRDVCRVF-AL-PFS		456
tr Q72ZD3 Q72ZD3_BACC1	VTLPDIDIDF	PDIRREDEMIRYVKDKYQGLRVAQIV	TFGTLAAKAAIRDARVM-GL-PER		436
tr Q97RC7 Q97RC7_STRPN	YTLPDIDIDF	PDIRYRDFIRYVGNKYGSKHAAQIV	TSTFGAKQALRDVLKRF-GV-PEY		402
tr Q4QMF6 Q4QMF6_HAEI8	VSMPDFDVDF	CMDGRDMVIDYVAEYGRGAVSQI	ITFGTMAAKAVIRDVGRVL-GH-PYG		454
tr Q2FG35 Q2FG35_STAA3	VTLPDIDIDF	EDTRRERVYQYVQEKYGEHVSGIVT	FGHLLARAVARDVGRIM-GF-DEV		400
tr A1R5G9 A1R5G9_ARTAT	VSMPDFDVDF	DDRRRSEVIDYVTKYGDERVAMIVTYG	TIKTQALKDSSRVL-GY-PFS		477
tr Q6NGE6 Q6NGE6_CORDI	PSAPDIDIDF	DDRRRGEMIRYAAQRWGEDKIAQVIT	FTGVTKTQALKDSARVQYQEGYK		476
tr Q6D8C8 Q6D8C8_ERWCT	VSMPDFDVDF	CMEKRDMDVIDYVAEYGRDAVSQI	ITFGTMAAKAVIRDVGRVL-GH-PYG		454
tr Q7NZL1 Q7NZL1_CHRVO	VSMPDFDVDF	CQENRWRVIEYTRRKYGEEAVSQIAT	FTGMTSSKSVIRDVGRVL-DL-PFG		456
tr Q5NM98 Q5NM98_ZYMMO	VSMPDFDIDF	CETRRSEVISYVQHKYGLDHVAQI	ITFGRMKARAVLKDTGRVL-QM-SYG		405
tr Q6KHF7 Q6KHF7_MYCMO	ITLPDIDIDF	QDDRRDEIVEYVTKYGHQKVALITTF	QRFQAKMALRDVGRNL-NF-TQV		382
tr C3K6H1 C3K6H1_PSEFS	VSMPDFDVDF	CMDGRDRVIDYVAEYGRNAVSQI	ITFGSMAAKAVIRDVARVQ-GH-PYG		456
tr B8ZR78 B8ZR78_MYCLB	TSLPDIDIDF	DDRRRGEMVRYAADKWGHRVAQVIT	FTGVIKTAKALKDSARIHYGQPGFA		472
tr A1F2E4 A1F2E4_VIBCH	VSMPDFDVDF	CMDKRDQVIDYVAEYGRDAVSQI	ITFGTMAAKAVIRDVGRVL-GH-PFG		459
tr C7C104 C7C104_HELPH	ISMPDFIDTF	CQRRRKEIIEYMIKYGKYNVAQVIT	FNKMLAKGVIRDVARVL-DM-PYK		478
tr Q2K9R0 Q2K9R0_RHIEC	VSMPDFDVDF	CQDRREIVRYVQAKYGRQVAQI	ITFGSLQARALRDVGRVL-EM-PYG		481
tr B5Y1I7 B5Y1I7_KLEP3	VSMPDFDVDF	CMEKRDQVIEHVADMYGRDAVSQI	ITFGTMAAKAVIRDVGRVL-GH-PYG		454
tr Q2KYR1 Q2KYR1_BORAI	VSMPDFDIDF	CQDNREIVRYVVKLYGRAAVSQIAT	FTGLGAKAVRDAGRVL-DM-PYM		459
tr Q5HV63 Q5HV63_CAMJR	VSMPDFIDVDF	CQDRRAEVIDYVIDKYGADKVAQVIT	FTGKLLAKGVIRDVARVC-DM-SIQ		479
	: * * * * :	* : : : : *	. : . : . : *		
tr Q8G7V8 Q8G7V8_BIFLO	TLGLVKMDFLGL	SNLTIHDTLKNIEANG-----K-PAIDYTKIPLDDAET	YKLMSR		630
tr C4ZRS6 C4ZRS6_ECOBW	YAGLVKDFDLGL	IRTLTIINWALEMINKRRAK----NGE-PPLDIAAIPLDDKKS	FDMQR		602
tr Q88MG5 Q88MG5_PSEPK	AAGLVKDFDLGL	IRTLTIIKWAMEIINREQAK----KNL-PDLNIDFIPLDDRKTY	ELLQK		604
tr C1DST1 C1DST1_AZOVD	AAGLVKDFDLGL	IRTLTIIKWAMETIHREQRR----RGETELVDIDRIALDDKATY	ALLQK		606
tr Q1D042 Q1D042_MYXXD	AAGLVKDFDLGL	IKTLTVIQHALDLVKNH-----G-KDIPRHEIPLNDEKT	WELMAK		610
tr Q72ZD3 Q72ZD3_BACC1	ELGLLKMDFLGL	IRNLTLLENIIKFIQKT-----G-KEIDIRNLPQDEKTF	QLLGR		579
tr Q97RC7 Q97RC7_STRPN	ASGLLKMDFLGL	IRNLTLFVQKMQELLAETE-----G-IHLKIEEIDLEDKET	LALFAS		542
tr Q4QMF6 Q4QMF6_HAEI8	YAGLVKDFDLGL	IRTLTIIKWALDIIINVRMVR----EGK-PRVDIAAIPLDDPES	FELLKR		602
tr Q2FG35 Q2FG35_STAA3	RIGLLKIDFLGL	IRNLSTIIHQILTQVKKDLGI-----NIDIEKIPFDDQKV	FELLQK		541
tr A1R5G9 A1R5G9_ARTAT	GLGLIKMDFLGL	IRNLTIISDALENIKMNR-----G-VLDLENLELDDAASY	ELLAR		631
tr Q6NGE6 Q6NGE6_CORDI	AIGLLKMDFLGL	IRNLTLVIGDCLENIKINR-----N-EELDLLEALQTD	DPNVKLLS		630
tr Q6D8C8 Q6D8C8_ERWCT	YAGLVKDFDLGL	IRTLTIIDWALEMINARRAK----LGQ-EPIDIAIPLGDKKS	FDMQR		602
tr Q7NZL1 Q7NZL1_CHRVO	QIGLVKDFDLGL	IRNLTIIEIAQYIKDLT-----G-EDVDVAHLPLDDKPA	YKVFAT		602
tr Q5NM98 Q5NM98_ZYMMO	KTGLVKDFDLGL	IKTSLVNLKAVQMLKAR-----G-VEVDLDHIPLDDAEV	FELLQK		550
tr Q6KHF7 Q6KHF7_MYCMO	NFSLKIDLLGL	IKALSTYIKLEQIKREAYKIKYFNFSNEDEY	TEQLKLDLPKKAFL		531
tr C3K6H1 C3K6H1_PSEFS	AAGLVKDFDLGL	IRTLTIIDWALKTINRDRAK----VNE-EPLDIAFIPLDDKPT	YSLLQK		604

```

472 tr|B8ZR78|B8ZR78_MYCLB AIGLLKMDFLGLRNLTIIGDAIENIKTNR-----G-IDLDLESVPLDDQATYELLGR 626
473 tr|A1F2E4|A1F2E4_VIBCH TAGLVKFDLGLRRLTLTIIDWALGLVNPRLLK----AGK-PPVRIEAIPLDDARSFRNLQD 607
474 tr|C7C104|C7C104_HELPHB PVDLIKFDLGLKTLTVIDDALKIKTQH-----N-IDVDFLSLDMDDPKVKYKTIQS 633
475 tr|Q2K9R0|Q2K9R0_RHIEC QAGLVKFDLGLKTLTVLVKVAADFVAKR-----G-INVDLAAIPLDDKKSFDFMLQR 626
476 tr|B5Y1I7|B5Y1I7_KLEP3 YAGLVKFDLGLRRLTLTIINWALEMINKRRAK----NGE-GPLDIAAIPLDDKKSFDFMLQR 602
477 tr|Q2KYR1|Q2KYR1_BORA1 AAGLVKFDLGLRNLTLILWAVRYVRQFN-----PDQ-RDFDIMALSLDDPAAYKVLCD 608
478 tr|Q5HV63|Q5HV63_CAMJR DVDLIKFDLGLKTLTVINNAIKLIKRY-----N-KDIIWETIDVNDISKVYKTIQS 625
479 .*:*** *: : : :
480
481
482 tr|Q8G7V8|Q8G7V8_BIFLO GDTLGVFQLDSDGMRSLKTLKPDNFNDISALIALYRPGEMSMDSHTNYAKRKNGLQKIT 690
483 tr|C4ZRS6|C4ZRS6_ECBOW SETTAVFQLESRGMKDLIKRLQPCDFEDMIALVALFRPGELQSGMVDNFIDRKHGREES 662
484 tr|Q88MG5|Q88MG5_PSEPK AETTAVFQLESRGMKELIKKLKPCDLEDLIALVALFRPGELQSGMVDDFINRKHGRAELA 664
485 tr|C1DST1|C1DST1_AZOVD AETTAVFQLESRGMKELIKKLKPDNIEDMIALVALFRPGELQSGMVDDFINRKHGRAELS 666
486 tr|Q1D042|Q1D042_MYXXD GDTAGIFQMESSGFTMVVKLPKNCDFEDVIAAGALYRPGELDSGMVDVFINRKHGREKVS 670
487 tr|Q72ZD3|Q72ZD3_BACC1 GDTTGVFQLESGGMRSLKLPNEFEDIVAVNSLYRPGEME--QIPTFIESKHGKRKIE 637
488 tr|Q97RC7|Q97RC7_STRPN GNTKGIFQFEQPGAIRLLKRVQPCFEDVVATSLNRPASD--YINNFVARKHQEEVT 600
489 tr|Q4QMF6|Q4QMF6_HAEI8 SETTAVFQLESRGMKDLIKRLQPCDFEDIALVALFRPGELQSGMVDNFIDRKHGREES 662
490 tr|Q2FG35|Q2FG35_STAA3 GDTTGIFQLESQDGRSVLKKLKEPFEDIVAVTSLYRPGEMEE--IPTYITRRHDPSKVQ 599
491 tr|A1F5E9|A1F5E9_ARTAT GDTLGVFQLESQDGRMRSLKLPNEFEDISAVLALYRPGEMGANAHDTYALRKHGIEVI 691
492 tr|Q6NGE6|Q6NGE6_CORDI GDTLGVFQLESQDGGMQELLKRMQPTGFNDIVASLALYRPGEMGVNAHWDYADRKNRKRPIE 690
493 tr|Q6D8C8|Q6D8C8_ERWCT SETTAVFQLESRGMKDLIKRLKPCDFEDMIALVALFRPGELQSGMVDNFIDRKHGREES 662
494 tr|Q7NZL1|Q7NZL1_CHRVO ANTTAVFQFESTGMKKMLVEAKPSKFEEIIAFVALYRPGEMD--LIPDFIQRMHGA-KFE 659
495 tr|Q5NM98|Q5NM98_ZYMMO GDTTGVFQLESQDGRMRSLKLPNEFEDIALVALYRPGEMDN--IPSGARKNGGEEIV 608
496 tr|Q6KHF7|Q6KHF7_MYCMO GNTAGVFQLESQPMQITIKVKKISKFSQDIVDVISLYRPGEMQY--IQEYIDNKNASKIK 589
497 tr|C3K6H1|C3K6H1_PSEFS AETTAVFQLESRGMKELIKKLKPCDLEDLIALVALFRPGELQSGMVDDFINRKHGRAELA 664
498 tr|B8ZR78|B8ZR78_MYCLB GDTLGVFQLDGGPMRDLRLRMQPTGFEDIVAVLALYRPGEMGMNAHNDYADRKNRQVIK 686
499 tr|A1F2E4|A1F2E4_VIBCH AKTTAVFQLESRGMKELIKRLQPCDFEDIALVALFRPGELQSGMVDNFIDRKHGREES 667
500 tr|C7C104|C7C104_HELPHB GDTVGFQIESGMFQGLNKLRLPSSFEDIIAIALYRPGEMESGMVDDFVNRKHGVEPIA 693
501 tr|Q2K9R0|Q2K9R0_RHIEC GETVGVFQVESAGMRKALIGMKPCIEDIIALVALYRPGEMEN--IPVYNARKHGGEEVE 684
502 tr|B5Y1I7|B5Y1I7_KLEP3 SETTAVFQLESRGMKDLIKRLQPCDFEDMIALVALFRPGELQSGMVDNFIDRKHGREES 662
503 tr|Q2KYR1|Q2KYR1_BORA1 ANTTAVFQLESRGMKELIKLRPNTFEDIIAMALYRPGELQSGMVDNFVNRKHGRAAD 668
504 tr|Q5HV63|Q5HV63_CAMJR GNTLGFQIESGGMQSLNARLKPERFEDIIAVLALYRPGEMESGMLDDFIDRKHGLKNIE 685
505 ..* :***: : : : * *** : : .
506
507
508 tr|Q8G7V8|Q8G7V8_BIFLO PIHPEL-EEPLKQVLDETYGLIIYQEQVQSAARILAGYSLGKADVLRRAMGKKKPEVLAK 749
509 tr|C4ZRS6|C4ZRS6_ECBOW YPDVQWQHESLKPVEPTYGIIYQEQVMQIAQVLSGYTLGGADMLRRAMGKKKPEEMAK 722
510 tr|Q88MG5|Q88MG5_PSEPK YPHSDYQYEGCLKPVLAPTYGIIYQEQVMQIAQVMAGYTLGGADMLRRAMGKKKPEEMAK 724
511 tr|C1DST1|C1DST1_AZOVD YPHPDYQYAGLEPKVLAPTYGIIYQEQVMQIAQVMAGYTLGGADMLRRAMGKKKPEEMAK 726
512 tr|Q1D042|Q1D042_MYXXD YPHPA-----LEPVLKDYTYGIIYQEQVMQISQVLGGYTLGRADLLRRAMGKKKAQVMA 725
513 tr|Q72ZD3|Q72ZD3_BACC1 YLH-----PDLKPILEPTYGIIYQEQVMQIASKLAFSLGADLLRRRAVSKKNRDLQ 692
514 tr|Q97RC7|Q97RC7_STRPN VLD-----PVLEDILAPTYGIIYQEQVMQVAQRLAGFSLGKADILRRAMGKKDASAMHE 655
515 tr|Q4QMF6|Q4QMF6_HAEI8 YPDAEYQHASLKPILEPTYGIIYQEQVMQIAQVLAGYTLGGADMLRRAMGKKKPEEMAK 722
516 tr|Q2FG35|Q2FG35_STAA3 YLH-----PHLEPILKNTYTYGIIYQEQVMQIASTFANFSYGEADILRRAMSKKNRAVLES 654
517 tr|A1R5G9|A1R5G9_ARTAT PIHPEL-EEPLKEILGGTFGLIYQEQVMQVAQKLAGYSLGQADILRRAMGKKKSELDK 750
518 tr|Q6NGE6|Q6NGE6_CORDI PIHPEL-EEPLKEILEETYGLIYQEQVMQIEISRLKANYTAGADGFRKAMGKKKPEVLAK 749
519 tr|Q6D8C8|Q6D8C8_ERWCT YPDIEWQHESLKPVEPTYGIIYQEQVMQIAQVLAGYTLGGADMLRRAMGKKNPVEMAK 722
520 tr|Q7NZL1|Q7NZL1_CHRVO YLH-----PLLEPVLAPTYGIIYQEQVMQSAQVIGGYSLGGADLLRRAMGKKKVEEMVA 714
521 tr|Q5NM98|Q5NM98_ZYMMO YPH-----PLLEPVLKETYGIIYQEQVMQAAQVLAGYSLGADLLRRAMGKKIQSEMDA 663
522 tr|Q6KHF7|Q6KHF7_MYCMO HVE-----KLYDEIVSKTHGIIYQEQVLEIVQKYAGLSYAKADILRRAMSKKNSEEIEK 644
523 tr|C3K6H1|C3K6H1_PSEFS YPHSDYQYEGCLKPVLAPTYGIIYQEQVMQIAQVMAGYTLGGADMLRRAMGKKKPEEMAK 724
524 tr|B8ZR78|B8ZR78_MYCLB PIHPEL-EEPLREILAETYGLIYQEQVMQIRIAQKVAGYSLARADILRRAMGKKKREVLEK 745
525 tr|A1F2E4|A1F2E4_VIBCH YPDEKWQHESLKEILEPTYGIIYQEQVMQIAQVLSGYTLGGADMLRRAMGKKKPEEMAK 727
526 tr|C7C104|C7C104_HELPHB YAF-----KELEPILKPTYTYGIIYQEQVMQIVQITIGGFSLGADLIRRAMGKKDAQIMAD 748
527 tr|Q2K9R0|Q2K9R0_RHIEC SIH-----PTIDHLLKETQGVIIYQEQVMQIAQVLSGYSLGADLLRRAMGKKIKAEMDQ 739
528 tr|B5Y1I7|B5Y1I7_KLEP3 YPDVQWQHESLKPVEPTYGIIYQEQVMQIAQVLSGYTLGGADMLRRAMGKKKPEEMAK 722
529 tr|Q2KYR1|Q2KYR1_BORA1 YFH-----PDLEGLTKSTYTYGIIYQEQVMQISQIIGGYSLGGADLLRRAMGKKKPEEMAK 723
530 tr|Q5HV63|Q5HV63_CAMJR YPF-----DSLEKVPLEPTYGIIYQEQVMQIVQIIGGFSLGADVVRAMGKKDPEKMKK 740
531 : * * :****: .. : . ** .*:*** :
532
533
534
535
536
537
538
539
540
541
542

```

```

543
544 tr|Q8G7V8|Q8G7V8_BIFLO EKVPFFAGMKEHGYSSEASQAVWDILVPFSGYAFNKAHSAAYGLISYWTAYLKTHYPVEF 809
545 tr|C4ZRS6|C4ZRS6_ECOBW QRSVFAEGAENGINAELAMKIFDLVEKFAGYGFNKAHSAAYALVSYQTLWLKAHYPAEF 782
546 tr|Q88MG5|Q88MG5_PSEPK QRGGFIEGCVANNIDADLAGNIFDLVEKFAGYGFNKAHSAAYGLVSYQTAWLKTHYPAPF 784
547 tr|C1DST1|C1DST1_AZOVD QRGGFVEGCAKNGIDAELAGNIFDLVEKFAGYGFNKAHSAAYGLVSYQTAWLKTRYPAF 786
548 tr|Q1D042|Q1D042_MYXXD ERAGFLEGCAKNNVDLKVAGEIFDLMEKFAEYGFNKAHSAAYGLVTIHTAWLKAHYPCEF 785
549 tr|Q72ZD3|Q72ZD3_BACC1 ERKHVFVQGCQNGYDETSAEKIYDLIVRFANYGFNRSHAVAYSMIGYQLAYLKANYTLEF 752
550 tr|Q97RC7|Q97RC7_STRPN MRASFIQGSLEAGHTVEKAEQVFDVMEKFAGYGFNKAHSAAYALVSYQTLWLKTHYPALF 715
551 tr|Q4QMF6|Q4QMF6_HAEI8 QRLVFKEGAENKNGVDGELSMKIFDLVEKFAGYGFNKAHSAAYALVSYQTLWLKTHYPALF 782
552 tr|Q2FG35|Q2FG35_STAA3 ERQHFIIEGAKQNGYHEDISKQIFDLILKFADYGFNKAHSAAYGLVSYWTAYLKANYTAEF 714
553 tr|A1R5G9|A1R5G9_ARTAT QFAGFSQGMQDNGYSMAAVKTLWDILLPFSFYAFNKAHSAAYGLVSYWTAYLKANYTAEF 810
554 tr|Q6NGE6|Q6NGE6_CORDI EYKKFSEGMFNSGYSKAAVDALWGTIEPFASYAFNKAHSAAYGLVSYWTAYLKANYTAEF 809
555 tr|Q6D8C8|Q6D8C8_ERWCT QRGGFEDGAKSRGVNGELAVKIFDLVEKFAGYGFNKAHSAAYALVSYQTLWLKAHYPAEF 782
556 tr|Q7NZL1|Q7NZL1_CHRVO QRAMFVQGAQKQDIPEAKANEIFDYMEKFAGYGFNKAHSAAYALVSYQTAWLKAHYPAEF 774
557 tr|Q5NM98|Q5NM98_ZYMO QRSRFVEGSKKHDITENQANQLFDLIDKFAGYGFNKAHSAAYALVSYQTAWLKAHYPAEF 723
558 tr|Q6KHF7|Q6KHF7_MYCMO FKDLFISSSIQHGQSEKVAEIEFSKIENFAQYGFNKAHSAAYALVSYQTAWLKAHYPAEF 704
559 tr|C3K6H1|C3K6H1_PSEFS QRGGFIEGCATNNIDADLAGNIFDLVEKFAGYGFNKAHSAAYGLVSYQTAWLKAHYPAEF 784
560 tr|B8ZR78|B8ZR78_MYCLB EFEGFSEGMQANGFSVHAIKALWDIILPFADYAFNKAHSAAYGLVSYWTAYLKANYTAEF 805
561 tr|A1F2E4|A1F2E4_VIBCH QRAVFQEGAENKNGVDGELAMKIFDLVEKFAGYGFNKAHSAAYALVSYQTLWLKTHYPALF 787
562 tr|C7C104|C7C104_HELPH NKAKFVEGAKNLGHDGQKAANLWDLIVKFAGYGFNKAHSAAYALVSYQTAWLKAHYPAEF 808
563 tr|Q2K9R0|Q2K9R0_RHIEC QRERFVEGAVKNGVKKPQADVIFDLLAKFANYGFNKAHSAAYALVSYQTAYMKAHYPVEF 799
564 tr|B5Y1I7|B5Y1I7_KLEP3 QRSIFEDGAKKNGVDGELAMKIFDLVEKFAGYGFNKAHSAAYALVSYQTLWLKAHYPAEF 782
565 tr|Q2KYR1|Q2KYR1_BORA1 HRELFOKGAVDKGHDPLAVKLFDLMEKFAEYGFNKAHSAAYALVSYQTAWLKAHYPAEF 783
566 tr|Q5HV63|Q5HV63_CAMJR LKTDADFADGAENKQYDRVKAEDLWELIVKFAGYGFNKAHSAAYALVSYQTAWLKAHYPAEF 800
567
568 * . . : : * : * * : : : :
569

```

Fig.2C CLUSTAL format for T-COFFEE Version of Polymerase III (alpha, catalytic subunit) (Prokaryotes)

The possible catalytic region is underlined and shaded yellow.

Only regions showing highly conserved blocks are shown.

Fig. 2D Prokaryotic DNA polymerases (pol IV)

sp Q8ZBZ9 DPO4_YERPE	M-----RKLIHVDMDCEFAAVEMRDDPRLRDIPLAIGGSKE---RRGVISTANY	46
sp Q99Y66 DPO4_STRP1	MLIFP-LINDTSRKLIHIDMDAFFAAVEERDNPAKLGKPVVIGKDPRETGGRGVVSTCNY	59
sp P63990 DPO4_SALTI	M-----RKLIHVDMDCEFAAVEMRDNPALRDIPIAIGGSRE---RRGVISTANY	46
sp P58964 DPO4_STAAW	M-----TERRLIHIDMDYFFAQVEMRDNPALKGKPVIVGGKAS---SRGVVSTASY	48
sp P63989 DPO4_SALTY	M-----RKLIHVDMDCEFAAVEMRDNPALRDIPIAIGGSRE---RRGVISTANY	46
sp Q5M2Y5 DPO4_STRT2	MLEFP-LINDTSRKLIHIDMDAFFAQVEMRDDPSLKDKPVIIGNDPRKTGGRGVVSTCNY	59
sp Q4ZWM4 DPO4_PSEU2	MT-----QRKLIHIDMDCEFAAIEMRDEPDLAGKPLAVGGSSE---RRGVISTCNY	48
sp A7Z6F3 DPO4_BACA2	MAG-----KGRIIFHIDMNSFYASVEMAYDPSLRGKPIAVAGNVKE--RKGIIVTCSY	51
sp Q9AK82 DPO4_STRCO	M-----RTAPTILHLDMDAFFASVEQASKPSLRGKAVVVGGLG---PRGVVATCSY	48
sp C3LQ59 DPO4_VIBCM	MQ-----DRIRKLIHVDMDCEFAAVEMRDNPAYREIALAVGGHEK---QRGVISTCNY	50
sp Q8X7Q1 DPO4_ECO57	M-----RKLIHVDMDCEFAAVEMRDNPALRDIPIAIGGSRE---RRGVISTANY	46
sp Q83M86 DPO4_SHIFL	M-----RKLIHVDMDCEFAAVEMRDNPALRDIPIAIGGSRE---RRGVISTANY	46
sp Q8XK37 DPO4_CLOPE	M-----KENRKLIHIDMDAFYASIEQRDNPKYKGLIVGGDPN---RRGVVATCSY	49
sp Q635E0 DPO4_BACCZ	MSTMREMPKPKGRVILHVDMNCFFASVEIAHDSSLQGKPLAVAGNEKE--RKGIIITCSY	58
sp Q88V07 DPO4_LACPL	M-----DTSRKLIHVDMDCEFAAIEEREHPAYKTQPLVIAHDPRQTGGRGVVSTANY	52
sp Q8NNP4 DPO4_CORGL	M-----QRWLHIDMDAFFASCEQLTRPTLRGRPVIVGGVS----GRGVVAGASY	46
sp Q9KP85 DPO4_VIBCH	MQ-----DRIRKLIHVDMDCEFAAVEMRDNPAYREIALAVGGHEK---QRGVISTCNY	50
sp Q6A763 DPO4_PROAC	M-----RSTASILHLDMDAFFASVEQRDKPSLVGKAVIVGGVG----GRGVVATASY	48
sp Q88NK4 DPO4_PSEPK	MS-----LRKLIHVDMDCEFAAIEMRDDPRLAGRPMVAGGSPD---HRGVISTANY	48
sp Q8DVR7 DPO4_STRMU	MLIFP-LINDTSRKLIHIDMDAFFAAVEERDNPKLGKPLVIGADPRQTGGRGVVSTANY	59
	* : : * : * : * . : : : : * : : : *	
sp Q8ZBZ9 DPO4_YERPE	PARRYGVRSAMPTAMAFK-LCPQLTLLPGRMAAYKEASQHIREIFARYTPLIEPLSLDEA	105
sp Q99Y66 DPO4_STRP1	EARKYGIHSAMSSKEAYE-RCPKAIFISGNYEKYRTVGDQIRRIKRYTDVVEPMSIDEA	118
sp P63990 DPO4_SALTI	PARQFVRSAMPTAMALK-LCPHLTLLPGRFDAYKEASRHVRDIFSRYSYSLIEPLSLDEA	105
sp P58964 DPO4_STAAW	EARKYGVHSAMPMSQAHK-LCPNGYFVTSNFGAYRETSQAQIMSIFRSYTDKVEPMSLDEA	107
sp P63989 DPO4_SALTY	PARQFVRSAMPTAMALK-LCPHLTLLPGRFDAYKEASRHVRDIFSRYSYSLIEPLSLDEA	105
sp Q5M2Y5 DPO4_STRT2	EARKYGVHSAMSSKEAYE-RCPNVAFISGNYSHYREVGMQIREIFKCYTDLVEPMSIDEA	118
sp Q4ZWM4 DPO4_PSEU2	EARAYGVRSAMSSRHALK-LCPDLTIVKPRMDAYKEASREIHTIFRDYTDLIEPLSLDEA	107
sp A7Z6F3 DPO4_BACA2	EARGVKTTPMVWRKR-LCPELIVVPPNFDYRSSSKEMFSILREYTDLVEPMSIDEA	110
sp Q9AK82 DPO4_STRCO	EARVFGVHSAMPMQARR-LCPHAAYLVPREFELYSISEQVMRLRLRELSPLVEPLSLDEA	107
sp C3LQ59 DPO4_VIBCM	QARKFVRSAMPTAQALK-LCPQLHVVPGRMSVYKSVSQQTIFQRYTSYSLIEPLSLDEA	109
sp Q8X7Q1 DPO4_ECO57	PARKFVRSAMPTGMALK-LCPHLTLLPGRFDAYKEASNHIREIFSRYSYSLIEPLSLDEA	105
sp Q83M86 DPO4_SHIFL	PARKFVRSAMPTGMALK-LCPHLTLLPGRFDAYKEASNHIREIFSRYSYSLIEPLSLDEA	105
sp P63989 DPO4_SALTY	EARKYGIHSAMPSLTAYK-LCPKAIIFRPRMEVYKVSQVMNILENSNIVEPLSLDEA	108
sp Q635E0 DPO4_BACCZ	EAREYGIRTTMPLWEAKR-LCPQLIVRRPNFTLYREASFQMFQILSRFTEKIQPVSIDEA	117
sp Q88V07 DPO4_LACPL	VARQFVHSAMPAAKALE-LCPTAVFKTPNFPLYREVSQAIFRIFHEYTEMIEPIAFIDEA	111
sp Q8NNP4 DPO4_CORGL	EARKFGARSAMPMHQAKARVGFAGVVTPRHIVYSAAARRVQIVEKRAGIVERLSIDEA	106
sp Q9KP85 DPO4_VIBCH	QARKFVRSAMPTAQALK-LCPQLHVVPGRMSVYKSVSQQTIFQRYTSYSLIEPLSLDEA	109
sp Q6A763 DPO4_PROAC	EARKFVHSAMAGSQARR-LAPNAAFSLGRFESYRESSKVMATLRELSPLVEPLSLDEA	107
sp Q88NK4 DPO4_PSEPK	EARAYGVRSAMSSRHALK-LCPDLLIVKPRFEAYREASREIHTIFRDYTELIEPLSLDEA	107
sp Q8DVR7 DPO4_STRMU	LAREFGIHSAMSSKEAYE-RCPQAIIFIRGNHTKYRIGLQVREIFRRYTDLVEPMSIDEA	118
	** * : : * . * . * . : : : : * : : : *	
sp Q8ZBZ9 DPO4_YERPE	YLDVSDSLAC---GGSATLIAQEIRQSIASE-LNLTASAGIAPIKFLAKIASELNKNP--	159
sp Q99Y66 DPO4_STRP1	YLDVTDNKLG---IKSAVKIAKLIQHDIWKE-VGLTCSAGVSYNKFIAKLASDFEKP--	172
sp P63990 DPO4_SALTI	WLDVTDSPHC---YGSATLIAREIRQTIFNE-LQLTASAGVAPVKFLAKIASDLNKNP--	159
sp P58964 DPO4_STAAW	YLDITELVRP---DLPASKIAQYIRKDILEQ-THLTASAGVSYNKFIAKLASGMNKP--	161
sp P63989 DPO4_SALTY	WLDVTDSPHC---YGSATLIAREIRQTIFNE-LQLTASAGVAPVKFLAKIASDLNKNP--	159
sp Q5M2Y5 DPO4_STRT2	YLDVTNKLK---IKSAVKVAKLIQYDIWQE-LHLTCSAGVSYNKFIAKLASDFQKPA--	172
sp Q4ZWM4 DPO4_PSEU2	FLDVSEAGHF---SGSATRIAQDIRRVSNO-LHITVSAGVAPNKFIAKIASDWKKPN--	161
sp A7Z6F3 DPO4_BACA2	YMDITDTPH---SHRAYETAEDIQARLQKE-LLLPSSIGIAPNKFIAKMASDMKKPL--	163
sp Q9AK82 DPO4_STRCO	FVDLDAGGAA-RDAETARLAGTKLRDITRV-TGLTGSVGLAASKMLAKIASAAKPD--	163
sp C3LQ59 DPO4_VIBCM	YLDVSESTAY---QGSATLIAQAIRDIWQE-LNLTASAGVAPIKFLAKVASDLNKNP--	163
sp Q8X7Q1 DPO4_ECO57	YLDVTDVSHC---HGSATLIAQEIRQTIFSE-LQLTASAGVTPVKFLAKIASDMNKNP--	159
sp Q83M86 DPO4_SHIFL	YLDVTDVSHC---HGSATLIAQEIRQTIFNE-LQLTASAGVAPVKFLAKIASDMNKNP--	159
sp Q8XK37 DPO4_CLOPE	FVDVSKSKRC---KGSATLIALEIKERIFKE-VGLTASAGVSFNKFIAKMASDFRKP--	162
sp Q635E0 DPO4_BACCZ	YLDITDCYA---LGSPLIAKMIQALLTE-LQLPCSIGIAPNLFIAKTASDMKKPL--	170
sp Q88V07 DPO4_LACPL	YLDVTNKKH---IHSAVELAHLRQQEIIHQ-THLTCTGISYNKFIAKLASDYRKPA--	165
sp Q8NNP4 DPO4_CORGL	FMEPEALVGA---TPEEVKQWAEELRAEIKEV-TGLPSSVGAGSGKQIAKIGSGEAKPD--	161
sp Q9KP85 DPO4_VIBCH	YLDVSESTAY---QGSATLIAQAIRDIWQE-LNLTASAGVAPIKFLAKVASDLNKNP--	163
sp Q6A763 DPO4_PROAC	FVDLEAGDVENLDALQQAVALRVARTEGLSCSVGIGSSKFMKAVASEMAKPD--	167
sp Q88NK4 DPO4_PSEPK	YLDVSDSQWY---SGSATRIAEDIRRVART-LHITVSAGVAPNKFIAKIASDWRKNP--	161
sp Q8DVR7 DPO4_STRMU	YLDVTENKLN---IKSAVKIAKLIQDIWEE-FHLTCSAGVSYNKFIAKLASDYDKPH--	172
	: : : . : : : : . * * : * : * *	

```

672
673 sp|Q8ZBZ9|DPO4_YERPE ---GQYVITPNQIQPFLQDLPLSKIPSVGAVTAKRLQALGLVTCGDIQKYPLAELLKHFG 216
674 sp|Q99Y66|DPO4_STRP1 ---GLTLVLKEDALCLAKLPIAKFHSVGKKS VKKLHDMGIYTGQDLLAVPEMTLIDHFG 229
675 sp|P63990|DPO4_SALTI ---GQYVITPADVPDFLKTLP LAKIPSVGKVSAAKLENMGLRTCGDIQQCDLAML LKRF 216
676 sp|P58964|DPO4_STAAW ---GMTVIDYQNVHDLMTLDIGDFPSVGKASKKVMHDNGIFNGRDLYEKTEFELIRLFG 218
677 sp|P63989|DPO4_SALTY ---GQYVITPADVPDFLKTLP LAKIPSVGKVSAAKLENMGLRTCGDIQQCDLAML LKRF 216
678 sp|Q5M2Y5|DPO4_STRT2 ---GLTVVLPBEAQEFLK LPIAKFHSVGKKSVERLHDMDIYTGADLLKISEITLIDRF 229
679 sp|Q4ZWM4|DPO4_PSEU2 ---GLFVITPDQVEEFVASLPVTKLHSGVKVTADKLGR LGIVDCADLR SRSKLALVREF 218
680 sp|A7Z6F3|DPO4_BACA2 ---GITILRKRELDPDLWPLPIEEMHGIGTKTAEKIKTLGIKTIGDLAKGDEHALKTL LG 220
681 sp|Q9AK82|DPO4_STRCO ---GLVLI PP GTERAMLEPMTVRTLP SVGPATGDHLRRAGITTVGEIAEAGEDELVRLLG 220
682 sp|C3LQ59|DPO4_VIBCM ---GLYVVTPDKVQEMVDSLPLEKIPSVGKVVALEKLHQAGLYVGADVRRADYRKLLHQFG 220
683 sp|Q8X7Q1|DPO4_ECO57 ---GQFVITPAEVSAFLQTLPLAKIPSVGKVSAAKLEAMGLR TC G DVQKCDLVILLKRF 216
684 sp|Q83M86|DPO4_SHIFL ---GQFVITPAEVPAPLQTLPLEKIPSVGKVSAAKLEAMGLR TC G DVQKCDLVILLKRF 216
685 sp|Q8XK37|DPO4_CLOPE ---GITVITEENSKDFIRKLP I GKFPSVGRVTKNKLNNIGVFKGEDLLGFSEKELIGILG 219
686 sp|Q635E0|DPO4_BACCZ ---GITVLRKRDIPEMIWPLPVGAMHGIGCKTAEKLN DIHIQTIEQLAKGNEHIIIRAKIG 227
687 sp|Q88V07|DPO4_LACPL ---GVTIVLPQDAEPFLLRPEIEKFP SVGKKTVPKMHDLGIKTGQDLYAQSEL DLIKQFG 222
688 sp|Q8NNP4|DPO4_CORGL ---GVFVVPVDKQHDLLDPLPVGALW SVGPVTGSKLASMGVETIGDLAALTQKEVEISLG 218
689 sp|Q9KPS5|DPO4_VIBCH ---GLYVVTPDKVQEMVDSLPLEKIPSVGKVVALEKLHQAGLYVGADVRRADYRKLLHQFG 220
690 sp|Q6A763|DPO4_PROAC HTDQIALIAPGTEADTIAPLSARAI PSVGPVTAERLEKLGRLTIADVRAARESELVHEL 227
691 sp|Q88NK4|DPO4_PSEPK ---GLFVITPNEVETFLVALPVARLH SVGKVTADKLTRLGIETCLHLREWSRLALVREF 218
692 sp|Q8DVR7|DPO4_STRMU ---GLTVILPQDAEGFLATLP I EKFYSVGKKSVEKLHALHIFTGKDVQQVPEMTLIDLF 229
693
694
695
696 sp|Q8ZBZ9|DPO4_YERPE RGVRLVGLHVTLLDPQL--ERQLLL----- 349
697 sp|Q99Y66|DPO4_STRP1 AGIRLLGV TMTNLEDKVA---DISL----- 362
698 sp|P63990|DPO4_SALTI RGVRLVGLHVTLLDPQL--ERQLVL----- 349
699 sp|P58964|DPO4_STAAW VPIRLIGVTGNLEQSTYK--NMTIY----- 353
700 sp|P63989|DPO4_SALTY RGVRLVGLHVTLLDPQL--ERQLVL----- 349
701 sp|Q5M2Y5|DPO4_STRT2 SGVRLLGVTTLTGLEDQEGR--QLDLD----- 364
702 sp|Q4ZWM4|DPO4_PSEU2 KPVRLLGIGVRLHDLRAA--HEQLELF----- 351
703 sp|A7Z6F3|DPO4_BACA2 NPVRLLGITGTELVEKQQAYKQLDLFSFKEDAKDEPIF----- 370
704 sp|Q9AK82|DPO4_STRCO GGVRLLGVGVSGLADYQTE--DLFAQAAGDRA--EEP AEEP GTEP-----AEAHSP 378
705 sp|C3LQ59|DPO4_VIBCM REIRLLGLSVMLKPELQ--MKQLSMFP----- 355
706 sp|Q8X7Q1|DPO4_ECO57 RGVRLVGLHVTLLDPQM--ERQLVL----- 349
707 sp|Q83M86|DPO4_SHIFL RGVRLVGLHVTLLDPQM--ERQLVL----- 349
708 sp|Q8XK37|DPO4_CLOPE KKVRLLGITISSLEENIITEEREQLS----- 356
709 sp|Q635E0|DPO4_BACCZ DSVRLLGVTATEIEWKTESVKQLDLFSFEEDAKEEPLL----- 377
710 sp|Q88V07|DPO4_LACPL SGRLLLGITLTGLAPLAFE--NLTLPL----- 361
711 sp|Q8NNP4|DPO4_CORGL GPIRLVGVSFSGLEESRQD--ILFPELDQQII--VPPAPD TDYEVGVQSSSSSESTQVEAP 386
712 sp|Q9KPS5|DPO4_VIBCH REIRLLGLSVMLKPELQ--MKQLSMFP----- 355
713 sp|Q6A763|DPO4_PROAC EGVRLLGVGVS NFATSAQE--ELFVVDDEGRLLDEPEV---TEE-----ITPIQP 383
714 sp|Q88NK4|DPO4_PSEPK KPVRLLGVGVRRLRDLRGA--HEQLELF----- 351
715 sp|Q8DVR7|DPO4_STRMU SGVRLLGVTVTALEDSTRE--ELSLTA----- 365
716
717
718

```

Fig. 2D CLUSTAL format for T-COFFEE Version of Polymerase IV (Family Y- Prokaryotes)

The possible catalytic region is underlined and shaded yellow.

Only regions showing highly conserved blocks are shown.

Fig. 2E Prokaryotic DNA polymerases (pol V- umuC, catalytic subunit)

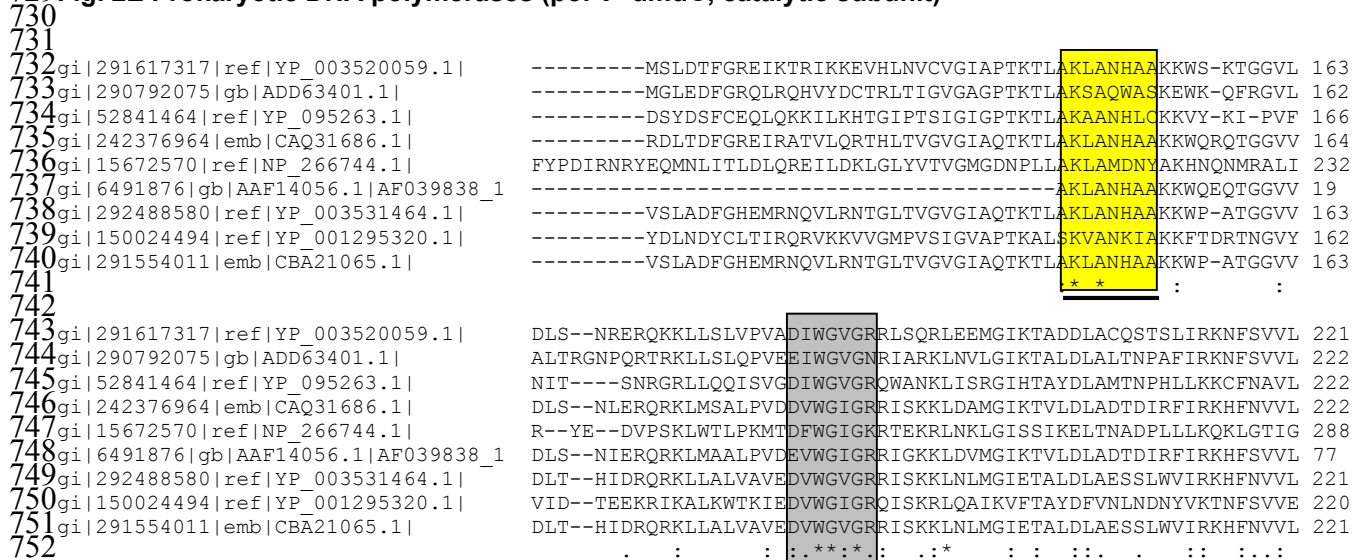


Fig. 2E CLUSTAL format for T-COFFEE Version of Polymerase V (Family Y- Prokaryotes)

The possible catalytic region is underlined and shaded yellow.
Only regions showing highly conserved blocks are shown.

Accession	Gene	Protein	Sequence	Length
sp P20311 DPOL_BPT3	RAAFGAEH-HLDGITGKPVWQAGIDASGLELRCLAHFMARFDN	-----GEYAHE	ILNG	503
sp P04415 DPOL_BPT4	KKMFAEEM-NAEAIKKIIMKGAGSCSTKPEVERYVKFSDDFLNE	-----LSNYTES	VLNS	537
sp P19822 DPOL_BPT5	VAPPGYRVIAWDLTTAEVYYAAVLSGD-RNMQQ-VFINMRNEPDKYPDFHSNIAHMF	-----GEYAHE	ILNG	547
sp P00581 DPOL_BPT7	RAAFGAEH-HLDGITGKPVWQAGIDASGLELRCLAHFMARFDN	-----GEYAHE	ILNG	503
	. . . *			
sp P20311 DPOL_BPT3	DIHTKNQMAAELPTRDNAK	TFIYGF	FLYAG	533
sp P04415 DPOL_BPT4	LIEECEKAATLANTNQLNR	KILINS	LYLAG	597
sp P19822 DPOL_BPT5	QCEPRDVKKLFPALRQA	AKAITF	GILYGS	577
sp P00581 DPOL_BPT7	DIHTKNQIAAELPTRDNAK	TFIYGF	FLYAG	533
	. :	..	.. ***.	
sp P20311 DPOL_BPT3	-DEKIGQIVG	-----	AGKERGKELK	552
sp P04415 DPOL_BPT4	INEYLNKVCGTNDEDFIAAGDTSVYVCVDKVI	EKGVLDRFKEQNDLVEFMNQFGKKKME		657
sp P19822 DPOL_BPT5	-PAKVAHSVNE	-----	ALLEQA	597
sp P00581 DPOL_BPT7	-DEKIGQIVG	-----	AGKERGKELK	552
	: : .		. :: :	
sp P20311 DPOL_BPT3	-----	-----	-----	KKF
sp P04415 DPOL_BPT4	PMIDVAYRELCDYMNREHLMHMDREAI	SCPPLGSKGVGGFWKAKKRYALNVYDMEDKRF		717
sp P19822 DPOL_BPT5	-----	-----	-----	EPF
sp P00581 DPOL_BPT7	-----	-----	-----	KKF
				:
sp P20311 DPOL_BPT3	LE-----	NTPA--IAALRESIQQT--	LVESSQWVAGEQQ	585
sp P04415 DPOL_BPT4	AEPHLKIMGMETQSSSTPKAVQEAL	ESIRRI--LQEGEESVQ-EYYKNFEKEYRQLDY		773
sp P19822 DPOL_BPT5	VE-----	CTV---ADAKEYIET	YFGQFPQLKRWIDKCHD	631
sp P00581 DPOL_BPT7	LE-----	NTPA--IAALRESIQQT--	LVESSQWVAGEQQ	585
	*	*	* *	

The possible catalytic region is underlined and shaded yellow. Only regions showing highly conserved blocks are shown.

801

802
803
804
805
806
807
808
809
810
811
812
813
814
815
816

• • •

821
822
823
824
825
826
827
828
829
830
831
832
833

• • • • •

837
838
839
840
841
842
843
844
845
846
847
848
849
850

11/11

854
855
856
857
858
859
860
861
862
863
864
865
866

	
---	---

```

871 sp|P33609|DPOLA_MOUSE LYPSIIQEYNICFTTVQRVSTSEVQKAT-----E-DEE--QEQIP----- 903
872 sp|P28040|DPOA_SCHPO LYPSIIQEYNICFTTVDRSPSNS-----DS--DDQIP----- 879
873 sp|P26019|DPOLA_DROME LYPSIIQEYNICFTTVQQPVDA-----DELPE----- 909
874 sp|P13382|DPOA_YEAST LYPSIIQEYNICFTTVDRNKE-----D-IDELP----- 894
875 sp|P09884|DPOLA_HUMAN LYPSIIQEYNICFTTVQRVASEAQKV-----E-DGE--QEQIP----- 899
876 sp|O89042|DPOLA_RAT LYPSIIQEYNICFTTVQRVASETLLKAT-----E-DEE--QEQIP----- 906
877 sp|Q9DE46|DPOLA_XENLA LYPSIIQEYNICFTTVHREAPSTGK-----GED--QDEIP----- 895
878 sp|Q9FHA3|DPOLA_ARATH LYPSIIQEYNICFTTVDRNSE-Q-----VF----- 945
879 sp|Q54SV8|DPOLA_DICDI LYPSIIQEYNVCFTTINRVKRDD-----GK--WEE-A----- 942
880 sp|O00874|DPOLA_LEIDO LYP SLIQEFNICFTTVDRSGSE-----IDVPPPENLICASCASCA 795
881 sp|P27727|DPOLA_TRYBB LYPSLIQEYNVCYTITDRDENTV-----S---AEVPPEPESLCLSCRA 793
882 sp|Q94636|DPOLA_OXYNO LYPSIIQEYNLCFTTVNRRTPTKNFDGSEMKNQYKKGENGEVEVDIEEA----- 912
883 sp|O48653|DPOLA_ORYSJ LYPSIIQEYNICFTTVDRSADGN-----VP----- 954
884 ****:*:*:*:*:*: : .
885
886
887
888 sp|P33609|DPOLA_MOUSE -ELPDPN--LEMGI LPREIRKLVERRRKQVKQLMKQQDLNPDLVLQYDIRQKA LKLTANSM 960
889 sp|P28040|DPOA_SCHPO -DTPSAS--ANQGIFPRRLIANLVERRRRIQIGLLKDNSATPTQRQWDIQQQA LKLTANSM 936
890 sp|P26019|DPOLA_DROME -TLPDSK--TEPGILPLQLRLVESRKVEKKLMAPDLSPELQMOMYHIROMA LKLTANSM 966
891 sp|P13382|DPOA_YEAST -SVPPSE--QTPGVLPRLLANLVDRRSREVKKVMKET--GLKYWELDIRQQA LKLTANSM 950
892 sp|P09884|DPOLA_HUMAN -ELPDPS--LEMGI LPREIRKLVERRRKQVKQLMKQQDLNPDLVLQYDIRQKA LKLTANSM 956
893 sp|O89042|DPOLA_RAT -ELPDPN--LDMGI LPREIRKLVERRRKQVKQLMKQQDLNPDLVLQYDIRQKA LKLTANSM 963
894 sp|Q9DE46|DPOLA_XENLA -ELPHSD--LEMGI LPREIRKLVERRRHVQLMKQPDNLNDLYLQYDIRQKA LKLTANSM 952
895 sp|Q9FHA3|DPOLA_ARATH -LRPSS--QTPGILPKLMEHLVSIRKS VKLMKKET--DHVLELDIRQQA LKLTANSM 999
896 sp|Q54SV8|DPOLA_DICDI -MPPPS--IEKGILPKVLHGLVSKRREIKRMEQEK-NKI IKAQYDIQQAA VKLIANSM 998
897 sp|O00874|DPOLA_LEIDO AGLSAPC--LHKCVLPKVIKSLVDSRREV KRLMKIEK-DANNLALLEIRQKA LKLTANSM 852
898 sp|P27727|DPOLA_TRYBB AGLPSPC--LHKCILPKVIRGLVDSRREIKRMMSKSEK-DPGNLAMLEIRQLA LKLTANSM 850
899 sp|Q94636|DPOLA_OXYNO -DLDPKNVNVLKDAVLPMLVLDLVQKRKAVDKMKTEK-DHVKLSQLEIRQQA LKLTANSM 970
900 sp|O48653|DPOLA_ORYSJ -NLPAS--KTGTGPLPELLKSLVERRRMVKS WLKTAS--GLKRQQFDIQQQA LKLTANSM 1008
901 . :*: **: *: :* : *.* **
902
903
904 sp|P33609|DPOLA_MOUSE YGCLGFYSRFYAKPLAALVTYKGREILMHTKMDMVQKM--NL E VIYGDTSIMINTNST- 1017
905 sp|P28040|DPOA_SCHPO YGCLGYTKSRFYARPLAVLITYKGREALMNTELADQM--GLQVIYGDTSVMINTNTV- 993
906 sp|P26019|DPOLA_DROME YGCLGFSAHSRFYAQH LAALVTHKGREILTNTQQLVQKM--NYDVVYGDTS LMINTNIT- 1023
907 sp|P13382|DPOA_YEAST YGCLGYYNSRFYAKPLAMLVTNKGREILMNT RQLAESM--NLVVYGDTS VMIDTGCD- 1007
908 sp|P09884|DPOLA_HUMAN YGCLGFYSRFYAKPLAALVTYKGREILMHTKEMVQKM--NL E VIYGDTSIMINTNST- 1013
909 sp|O89042|DPOLA_RAT YGCLGFYSRFYAKPLAALVTYKGREILMHTKEMVQKM--NL E VIYGDTSIMINTNST- 1020
910 sp|Q9DE46|DPOLA_XENLA YGCLGFYSRFYAKPLAALVTYKGREILLHTKEMVQKM--NL E VIYGDTSIMINTNCN- 1009
911 sp|Q9FHA3|DPOLA_ARATH YGCLGFENS RFYAKPLAELITLGGRDILQRTVDLVQN-HLNLEVIYGDTSIMIHSGLD- 1057
912 sp|Q54SV8|DPOLA_DICDI YGCLGFSSH RFYALPLAELVTRKGRLENLQ GASIVNKM--CYDVIYGDTS LMIYTGVG- 1055
913 sp|O00874|DPOLA_LEIDO YGCLGF EYSRFYAHQPLAELVTRQGRALQSTVDLIPQLNPSLRVIYGDTSVMIMQTGIKN 912
914 sp|P27727|DPOLA_TRYBB YGCLGF EYSRFYAHQPLAELVTRQGRALQNTVELIPQISPIRVYGDTSVMIMQTGIKR 910
915 sp|Q94636|DPOLA_OXYNO YGCLGFESSRFYAHQAATAALITRTGRETLLRTKDIAEN-KLGFNVVYGDTSIMINTGSN- 1028
916 sp|O48653|DPOLA_ORYSJ YGCLGF S N SRFYAKPLAELITLGGREILQNTVDLVQN-NLNLEVIYGDTSIMIH TGLD- 1066
917 *****: ***.* : *:* * * * . : . *.*****:*. :.
918
919
920 sp|P33609|DPOLA_MOUSE NLEE VFK LG NKVK SEVNKLYK--LLEID IDAV FKS LLLL KKKY AALVVEPT-SDGN YIT 1074
921 sp|P28040|DPOA_SCHPO DKNHALRIGNEFK EKVNERYS--KLEID ID NVY QRM LL HAK KKY AALQ L DS Q---G--KP 1046
922 sp|P26019|DPOLA_DROME DYDQVYKIGHNIKQS VNKLYK--KLEID ID G VF GC LLLL KKKY AA IKLSK--SKGN LR 1080
923 sp|P13382|DPOA_YEAST NYADA IKIGL GF KR LVNER YR--LLEID ID NV F KK LLL HAK KKY AALT VN LD-KNG NG TT 1064
924 sp|P09884|DPOLA_HUMAN NLEE VFK LG NKVK SEVNKLYK--LLEID ID G VF KS LLLL KKKY AALVVEPT-SDGN YVT 1070
925 sp|O89042|DPOLA_RAT NLEE VFK LG NKVK NEVNKLYK--LLEID ID G VF KS LLLL KKKY AALVVEPT-SDGN YIT 1077
926 sp|Q9DE46|DPOLA_XENLA NLEE VFK LG NR VK SEIN KSYK--LLEID ID G IF KS LLLL KKKY AALS VT E PT-GDG KYVT 1066
927 sp|Q9FHA3|DPOLA_ARATH DIEEV KA IKS KV IQ EVN KKYR--CLK ID CD GI YKR M LL R KKKY AV KLQ-F-KDG KPCE 1113
928 sp|Q54SV8|DPOLA_DICDI TFNEA ET IG KEI QK K IN D QY RG SV ME I GL DG I FK RL LL F KKKY AC LKE F RI-DST TTKC 1114
929 sp|O00874|DPOLA_LEIDO DIKAV RD LG DL K AK IN KRYQ--SLEID ID G V FR AI LL L KKKY ALT VT DW-QGEG KTY 969
930 sp|P27727|DPOLA_TRYBB DIVK RV NL G LF IK GK VN KRYQ--SLEID ID G V FR AM LL L KKKY ALS VDW-QGEG KYV 967
931 sp|Q94636|DPOLA_OXYNO QLQQA LE MG K RL KG EV NC LYK--CLEI E ID G V FS LLLL KKKY AL KY EN FL SP AE VKV 1086
932 sp|O48653|DPOLA_ORYSJ DISRA KGI AG VI QE VN KKYR--CLEI DL DG I YKR M LL L KKKY AI KVA-L--DGSL RE 1121
933 . : . * * : : * : : * * *****:
934
935
936
937
938
939
940
941

```

```

942 sp|P33609|DPOLA_MOUSE KQELKGLDI VRRDWCDLAKDTGNFVIGQILSDQSRDTIVENIQKRLIEIGENVLNGSV-P 1133
943 sp|P28040|DPOA_SCHPO NLDVKGGLDMKRREFCTLAKEASKFCLDQILSGELTETVNIHSYLMDFSEKMRNGKF-P 1105
944 sp|P26019|DPOLA_DROME EQEHKGLDI VRRDWSQLAVMVGKAVLDEVLSEKPLEEKLDVHAQLEKIKTQIAEGVV-P 1139
945 sp|P13382|DPOA_YEAST VLEVKGGLDMKRREFCPLSRDVSIIHVLNTILSDKDPPEALQEVYDYLEDIRIKVETNNI-R 1123
946 sp|P09884|DPOLA_HUMAN KQELKGLDI VRRDWCDLAKDTGNFVIGQILSDQSRDTIVENIQKRLIEIGENVLNGSV-P 1129
947 sp|O89042|DPOLA_RAT KQELKGLDI VRRDWCDLAKDTGNFVIGQILSDQSRDTIVENIQKRLIEIGENVLNGSV-P 1136
948 sp|Q9DE46|DPOLA_XENLA KQELKGLDI VRRDWCELAKQAGNYVISQILSDQPRDSIVENIQKKLTEIGENVVTNGTV-P 1125
949 sp|Q9FHA3|DPOLA_ARATH DIERKGVDIVRRDWSLLSKEIGDLCLSKILYGGSCEDVVEAIHNELMKIKEEMRNGQV-A 1172
950 sp|Q54SV8|DPOLA_DICDI ERENKGIDIVRRDYCDLTKDIGQWVLNLILGGEEKIALFSLIKEYLESVQQIKDNTL-A 1173
951 sp|O00874|DPOLA_LEIDO KKEVGGLDMVRRDWCPLSKVCDSVLSRVLNAEGSEDILDYVMNYMRDVSEKVRAGRY-T 1028
952 sp|P27727|DPOLA_TRYBB KREVKGGLDMVRRDWCPLSQHVSDAVLKRIINAEGGEDILDFVIKYMKGVAQDVRSGNVYP 1027
953 sp|Q94636|DPOLA_OXYNO VQEMKGGLDMVRRDWCPLSKRVGRVLDQILSGKQREEVVLNLNEFLSNIGNELKEGTI-K 1145
954 sp|O48653|DPOLA_ORYSJ NIERKGGLDMVRRDWSLLSKEIGDFCLNQILSGGSCDDVIESIHSSLVQVQEQMRGGQT-E 1180
955
956 : ***: ***:.. *: : : * . : : . .: .
957
958
959
960
961
962
963
964
965
966
967
968
969
970

```

Fig. 2G CLUSTAL format for T-COFFEE Version of Alpha polymerases (Eukaryotes)
The possible catalytic region is underlined and shaded yellow.
Only regions showing highly conserved blocks are shown.

Fig. 2H Beta polymerases (Eukaryotes)

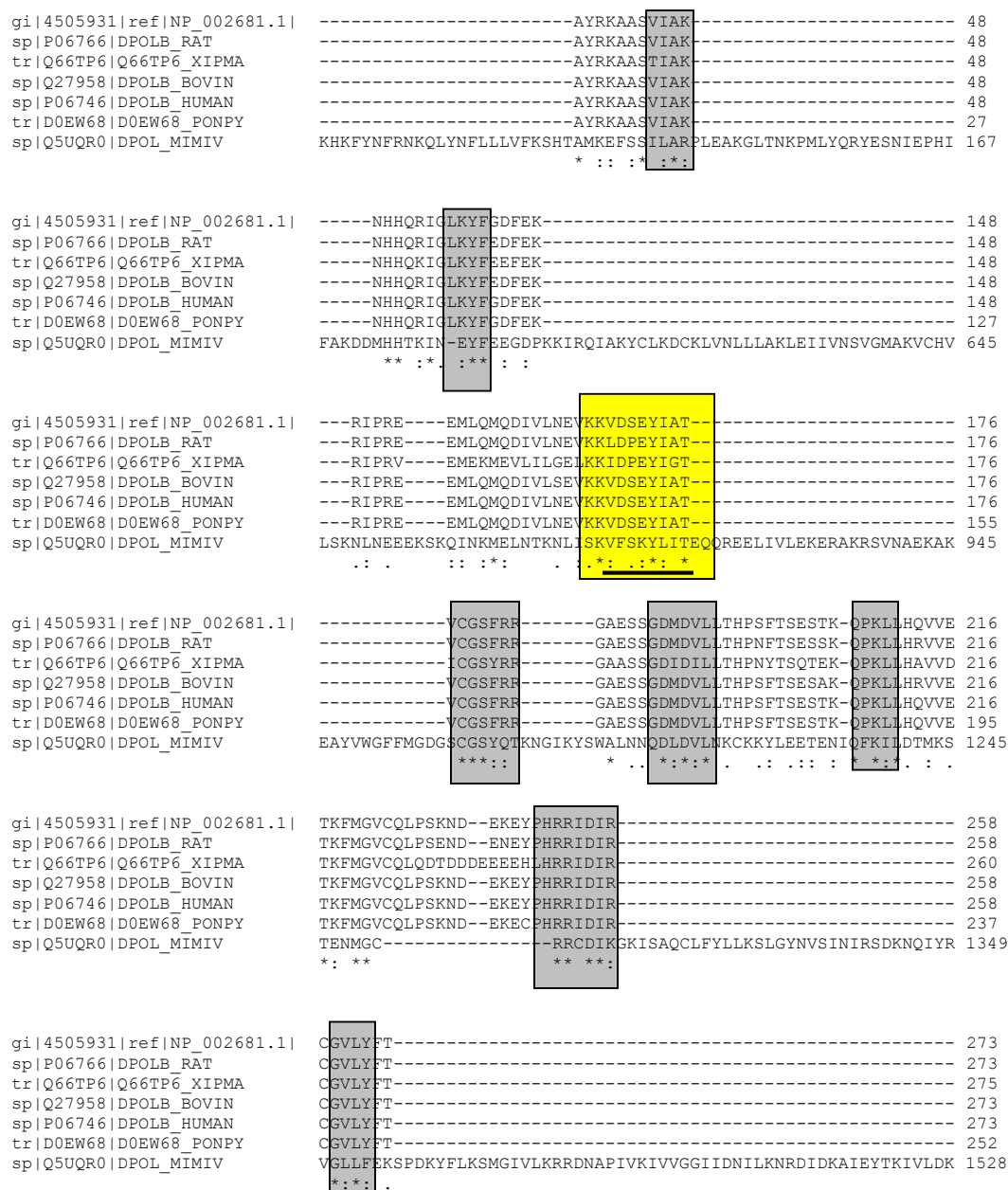


Fig. 2H CLUSTAL format for T-COFFEE Version of Beta Polymerases (Eukaryotes)

The possible catalytic region is underlined and shaded yellow.

Only regions showing highly conserved blocks are shown.

Fig. 2I Gamma polymerases (Eukaryotes, Mitochondrial)

sp Q01941 DPOG_PICPA	-----IRAEDLNKFTFVKVPHFDGPSARVTNCMTKSCLGFF	567
sp Q12704 DPOG_SCHPO	-----DTKLDYNNYIFFKVPHKDGPPEARCGSPLSKSYQRYF	574
sp P15801 DPOG_YEAST	-----NLGLQC-TGVLFKVPHPNGPTFNCTNLLTKSYNHFF	576
sp Q9Y767 DPOG_NEUCR	-----RLRMDV-DHKYFKLPHKDGPNARCVPNMAKGYLPYF	626
sp Q92076 DPOG1_CHICK	EDGLPE-LVEESSQPSFHHNGPY-NDVNIPGCWFFKLPHKDGNENNNGVSPFAKDFLPRM	186
sp P54098 DPOG1_HUMAN	PLALTARGGPKDTQPSYHHNGPY-NDVDIPGCWFFKLPHKDGNNSCNVSGSPFAKDFLPKM	774
sp Q27607 DPOG1_DROME	YKKL---SQKQQRLETQYQSGVWCNKVLDCCFFLKLPHKNGPSPFRVGNPLSKDFLNKF	693
sp P54099 DPOG1_MOUSE	PLVLPAACAPKSSQPTYHHNGPY-NDVNIPGCWFFKLPHKDGNNNYVSGSPFAKDFLPKM	753
sp Q91684 DPOG1_XENLA	PVKL---EMEFDSLDPNHHGNSPC-GDVNVSGCWFFKLPHKDGNANNVSGSPFAKDFLPKM	744
	*** : * . . : * :	
sp Q01941 DPOG_PICPA	EKGFLNSQYP---LAKDALQMAVASSYWTSSRERIMNQFVVFE-----	607
sp Q12704 DPOG_SCHPO	EEGILQSDYE---VAKKALEMSASCYWSSARDRISSQMVVWDKDAELG-V-----P	622
sp P15801 DPOG_YEAST	EKGVLKSESE---LAHQALQINSSGSYWMSARERIQSQFVVPSCFPNEFQSLSAKSSLN	633
sp Q9Y767 DPOG_NEUCR	EKGILSSEYP---YAKEALEMNASCYWISARERIKNQMVVYEDQLPPS-QRFVNKD-AD	681
sp Q92076 DPOG1_CHICK	EDGTLRAAVG-RTHGTRALEINKMVSFWRNAHKRVSSQVVVWLKKGELP-RAVTRHP-AY	243
sp P54098 DPOG1_HUMAN	EDGTLQAGPG-GASGPRALEINKMISFWRNAHKRISSQMVVWLPRSALP-RAVIRHP-DY	831
sp Q27607 DPOG1_DROME	AENVLSGGDPSQAAARVIDIARMMSYWRNDRIMGQMVVWLDQQQLP-NEFTGEK---	749
sp P54099 DPOG1_MOUSE	EDGTLQAGPG-GASGPRALEINKMISFWRNAHKRISSQMVVWLPRSALP-RVVTTRHP-AF	810
sp Q91684 DPOG1_XENLA	EDGTLQASTG-DSSATRALEINKMISFWRNAHKRISSQMVVMMKKNELH-RTITTRDP-EF	801
	. . * : . . : : * * . . : * : . * . *	
sp Q01941 DPOG_PICPA	----DDMGYILPQIIPMGTTITTRAVENTWLTASNAKKNRLGSELHSLIEAPKGYCFVGAD	663
sp Q12704 DPOG_SCHPO	S-SVDGFGIILPCIIPMGTVTTRAVENTWLTASNKKNRIGSELHAMIRAPDGYTFVGAD	681
sp P15801 DPOG_YEAST	NEKTNDLAIILPKIVPMGTTITTRAVENTWLTASNKANRIGSELHTQVKAPPGYCFVGAD	693
sp Q9Y767 DPOG_NEUCR	SN-TPIGGFVLPQVIMPGTTITTRAVENTWLTASNKKNRVGSSELHAMVRAPPGYCFVGAD	740
sp Q92076 DPOG1_CHICK	SE-EEDYGAILPQVVTAAGTTITTRAVENTWLTASNARADRVGSELHAMVQPPPGYSLVGAD	302
sp P54098 DPOG1_HUMAN	DE-EGLYGAILPQVVTAAGTTITTRAVENTWLTASNARPDVGSSELHAMVQAPPGYTLVGAD	890
sp Q27607 DPOG1_DROME	CQ-PIAYGAICPQVVACGTLTTRAMEPTWMTASNRPDRLGSELHSMVQAPPGYTLVGAD	808
sp P54099 DPOG1_MOUSE	DE-EGHYGAILPQVVTAAGTTITTRAVENTWLTASNARPDVGSSELHAMVQAPPGYTLVGAD	869
sp Q91684 DPOG1_XENLA	DE-ENKYGAILAQVVSAGTTITTRAVENTWLTASNARADRVGSELHAMVQPPPGYHLIGAD	860
	. : . . : . * * * * * : * * * * : : * * * * : . . * * * * *	
sp Q01941 DPOG_PICPA	VDSEELWIAVLIGDSVF-KIHGGTAIGWMTLEGTKNEGTDLHSKTAKILGISRNEAKIFN	722
sp Q12704 DPOG_SCHPO	VDSEELWIAVLMGDSQF-RLHGATALGMMTLEGKKSEGTDLHSKTAAILGVSRDSAKVFN	740
sp P15801 DPOG_YEAST	VDSEELWIASLVGDSIF-NVHGGAIGWMCLEGTKNEGTDLHTKTAQILGCSRNEAKIFN	752
sp Q9Y767 DPOG_NEUCR	VDSEELWIASVVGDATAF-KLHGGAIGFMTLEGTKSQGTDLHSRTASILGITRNDAKVFN	799
sp Q92076 DPOG1_CHICK	VDSQELWIAAVLGEAHFAGMHGCTAFGWMTLQGGKSGTDLHSKTAAATVGISREHAKVFN	362
sp P54098 DPOG1_HUMAN	VDSQELWIAAVLGDAAHFAGMHGCTAFGWMTLQGRKSRGTDLHSKTATTVGISREHAKIFN	950
sp Q27607 DPOG1_DROME	VDSQELWIASVLGDAAFCAGHGCTPLGWMTLSGSKNSGSDMHSITAKAVGISRDHAKVIN	868
sp P54099 DPOG1_MOUSE	VDSQELWIAAVLGDAAHFAGMHGCTAFGWMTLQGRKSRGTDLHSKTAAATVGISREHAKIFN	929
sp Q91684 DPOG1_XENLA	VDSQELWIAAILGEAHFAGIHGCTAFGWMTLQGGKSSGTDLHSKTASTVGISREHAKVFN	920
	*** * * * : : : : * * . . : * * * * . * * : : * * : * : : * * : *	
sp Q01941 DPOG_PICPA	YGRIYGAGIKFTTTLLKKFNPALSDAEAKATANALYTATKGIS-GR-----	767
sp Q12704 DPOG_SCHPO	YGRLYGAGLKHTTTLLMQMNPTLKTAEAKELAKKLYASTKGYSKM-----	786
sp P15801 DPOG_YEAST	YGRIYGAGAKFASQLLRKFNPSLTDEETKKIANKLYENTKGKT-KR-----	797
sp Q9Y767 DPOG_NEUCR	YGRIYGAGLKFAASQLLRQFNPSLTAEATTAIATKLYDATKGAKTNR-----	845
sp Q92076 DPOG1_CHICK	YGRIYGAGQPFAERLLMQFNHRLTQQQAREKAQQMYAVTKGIRRFHLSEEGEWLVKELEL	422
sp P54098 DPOG1_HUMAN	YGRIYGAGQPFAERLLMQFNHRLTQQEAAEKAQQMYAATKGLRWYRLSDEGEWLVRELN	1010
sp Q27607 DPOG1_DROME	YARIYGAGQLFAETLLRQFNPTFSASEAKAKAMKFSITKGRVYRLREEFHDELD---	925
sp P54099 DPOG1_MOUSE	YGRIYGAGQSFAERLLMQFNHRLTQQEAAEKAQQMYAVTKGLRRYRLSADGEWLVKQLNL	989
sp Q91684 DPOG1_XENLA	YGRIYGAGQPFAERLLMQFNHRLTQQEAAEKAQQMYAVTKGIRRYILSKEGEWLVVEELGI	980
	* . * * * * : . : * * : * : . : : * : : * * *	
sp Q01941 DPOG_PICPA	-----Y-----DKKSIWYGGSSEIIINRLEAIAEMAH	794
sp Q12704 DPOG_SCHPO	-----SK-RLQEMGLPKLTFWSQGTESFVFNKLEAMAQLPS	821
sp P15801 DPOG_YEAST	-----SK-----LFKKFWYGGSESILFNKLESIAEQET	825
sp Q9Y767 DPOG_NEUCR	-----KS-----LYKRPFWRGGTESFVFNMLEFAEQER	874
sp Q92076 DPOG1_CHICK	AVDKAEDGTVSAQDVQKIQREAMRKSRRKKKWDVVAHRMWAGGTSEMFNKLSEALSAS	482
sp P54098 DPOG1_HUMAN	PVDRTEGGWISLQDLRKVQRETARKSQW-KKWEVVAERAWKGGTESEMFNKLSEIATSDI	1069
sp Q27607 DPOG1_DROME	-----RAYSSYE-----SRLAIQRN--RTLAEVFHRPNWQGGTESAMFNRLSEIATGSQ	973
sp P54099 DPOG1_MOUSE	PVDRTEGDGWSLQDLRMIRREASRKSrw-KKWEVVAERAWTGGTESEMFNKLSEIAMSdt	1048
sp Q91684 DPOG1_XENLA	SVERGEENSvNLQDLRKIQDKDATKRSR--RKWNLVSRRIWTGGTESQMFNKLSEIAMSps	1038
	* * * * : * * * : *	

Only regions showing highly conserved blocks are shown.

Fig. 2J Delta polymerases (Eukaryotes)

1171	sp P30316 DPOD_SCHPO	PYFYVKAPVGFRPEM---LERFTQDLDATCN-----G---GVIDHCIIEMK	176
1172	sp P15436 DPOD_YEAST	NYLYVPAPNSSDANDQEQINKFVHYLNETF-----D---HAIDSIEVVSK	199
1173	sp P28339 DPOD1_BOVIN	PYFYTPAPPFGFGEH---LSELQRELSAAISRDRG-G---KELTG---PAVLAVELCSR	194
1174	sp P28340 DPOD1_HUMAN	PYFYTPAPPFGFGEH---MGDLQRELNLAISRDRG-G---RELTG---PAVLAVELCSR	195
1175	sp O54747 DPOD1_RAT	PYFYTPAPPFGFAEH---LSELQRELNAAISRDRG-G---KELSG---PAVLAVELCSR	191
1176	sp P30315 DPOD1_PLAFK	PYFYVEKPPDDFDNED---LIKLEMLNENLNLSQY-----KIYE---KKILKIEIVKT	187
1177	sp P54358 DPOD1_DROME	PYFYIEAPSQFEEHH---CEKLQKALDQKVIADIRN-N---KDNVQ---EAVLMVELVEK	175
1178	sp P52431 DPOD1_MOUSE	PYFYTPAPPFGFAEH---LSELQRELNAISRDRG-G---KELSG---PAVLAVELCSR	193
1179	sp P90829 DPOD1_CAEEL	PHFYFQAPQGFGEH---IGTAQSAICNMVAAKRRGGSGQAQLPGKVVDNLVHVEIVHG	163
1180	sp Q9LVN7 DPOD1_ARATH	PYFYIACPPGMGPDD---ISNFHQSLGRMRESNKN-A-----KVP---KFVKRIEMVQK	190
1181	sp Q9LRE6 DPOD1_ORYSJ	PYFYISCPMGMPDD---ISRFHQTLGRMKDSNRN-S-----NVP---RFVKRIELVQK	200
1182	sp P46588 DPOD_CANAL	HYFYCPVPKGFE-EN---LTEFTNYLKATF-----DGIERVEITSK	136
1183	sp P97283 DPOD1_MESAU	PYFYTPAPPFGFAEH---LSDLQRELSAISRDQRG-G---KELSG---PAVLAVELCSR	191
1184	sp O48901 DPOD1_SOYBN	PYFYICCPMGMPDD---ISHFHQTLGRMREANRN-S-----NVG---KFVRRIEMVQR	157
1185		::* * :	
1186			
1187			
1188			
1189			
1190			
1191	sp P30316 DPOD_SCHPO	APLRIMSFSDIECAGRKGV-FPDPSIDPVIQIASIVTQYGDST-PFVRNVFCVDTCSQIVG	349
1192	sp P15436 DPOD_YEAST	APLRIMSFSDIECAGRIGV-FPEPEYDPVIQIANVVSIAGAKK-PFIRNVFTLNTCSPITG	370
1193	sp P28339 DPOD1_BOVIN	APLRVLSFSDIECAGRKGI-FPEPERDPVIQICSLGLRWGEPE-PFLRLALTLPFCAPILG	364
1194	sp P28340 DPOD1_HUMAN	APLRVLSFSDIECAGRKGI-FPEPERDPVIQICSLGLRWGEPE-PFLRLALTLPFCAPILG	365
1195	sp O54747 DPOD1_RAT	APLRVLSFSDIECAGRKGI-FPEPERDPVIQICSLGLRWGEPE-PFLRLALTLPFCAPILG	361
1196	sp P30315 DPOD1_PLAFK	PKLRILSFSDIECIKLDGKGFPFAKNDPIIQISSILYFQGEFIDNCTKFIPTLLECASIPG	359
1197	sp P54358 DPOD1_DROME	APFRILSFSDIECAGRKGI-FPEAKIDPVIQIANMVIRQGERE-PFIRNVFTLNECAPIIG	349
1198	sp P52431 DPOD1_MOUSE	APLRVLSFSDIECAGRKGI-FPEPERDPVIQICSLGLRWGEPE-PFLRLALTLPFCAPILG	363
1199	sp P90829 DPOD1_CAEEL	APIRILSFSDIECIGRRGV-FPEAIDPIIQIANLVKIEGEAE-PFVRNCFVLGTCPAVVG	335
1200	sp Q9LVN7 DPOD1_ARATH	APFRVLSFSDIECAGRKGH-FPEAKHDPVIQIANLVTLQGEDH-PFVRNVMTLKSAPIVG	359
1201	sp Q9LRE6 DPOD1_ORYSJ	APFRILSFSDIECAGRKGH-FPEPTHDPVIQIANLVTLQGEQG-PFVRNVMTLKSAPIVG	369
1202	sp P46588 DPOD_CANAL	APLRILSFSDIECAGRKGV-FPEAEHDPVIQIANVVQKSGESK-PFVRNVFTVNTCSSIIG	299
1203	sp P97283 DPOD1_MESAU	APLRVLSFSDIECAGRKGI-FPEPERDPVIQICSLGLRWGEPE-PFLRLALTLPFCAPILG	361
1204	sp O48901 DPOD1_SOYBN	APFRILSFSDIECAGRKGH-FPEPTHDPVIQIANLVTLQGEDQ-PFIRNVMTLKSAPIVG	326
1205		. :* ::**** * **: **::**::* : : : *:: *	
1206			
1207			
1208			
1209	sp P30316 DPOD_SCHPO	TQVYEFQNAEMLSSWSKFVRDVPDVLIIGYNICNFDIPYILDRAKSLRIHNFPLLGRIH	409
1210	sp P15436 DPOD_YEAST	SMIFSHATEEEMLSNWRNFIIKVPDVIIGYNTTNFDIPYILNRAKALKVNDFFYFGRLK	430
1211	sp P28339 DPOD1_BOVIN	AKVQSYEREEDLLQAWSTFIRIMDPDVIITGYNIQNFDPYIISRAQTLKVPGFPLGRVI	424
1212	sp P28340 DPOD1_HUMAN	AKVQSYEKEEDLLQAWSTFIRIMDPDVIITGYNIQNFDPYIISRAQTLKVQTFPFLGRVA	425
1213	sp O54747 DPOD1_RAT	AKVQSYEREEDLLQAWATFILAMPDVIITGYNIQNFDPYIISRAQTLKVDRFPFLGRVT	421
1214	sp P30315 DPOD1_PLAFK	SNVIWFNDEKTLLEAWNEFIIRDPDVLITGYNIINFDLPYILNRTALNLKKLFLGRIK	419
1215	sp P54358 DPOD1_DROME	SQVLCHDKETQMLDKWSAFVREVPDVLITGYNIINFDLPYILNRAHLKVRNFPFLGRIK	409
1216	sp P52431 DPOD1_MOUSE	AKVQSYEREEDLLQAWADFILAMPDVIITGYNIQNFDPYIISRAQALKVDRFPFLGRVT	423
1217	sp P90829 DPOD1_CAEEL	SNIIQCVNEKVLLEKWAEFVREVPDVIITGYNINFDLPYILDRAKVLSLPQVSHLGRQK	395
1218	sp Q9LVN7 DPOD1_ARATH	VDVMSFETEREVLLAWRDILRVPDVIITGYNICKFDLPYIIERAATLGIEEFPLGRVK	419
1219	sp Q9LRE6 DPOD1_ORYSJ	VDVMSFETEREVLLAWRDILRVPDVIITGYNICKFDLPYIIERAEVLKIVEFPILGRIR	429
1220	sp P46588 DPOD_CANAL	SQIFEHQREEDMLMHWEKITKVPDVIIGYNTANFDIPYILNRAKALGLNDFPFFGRLK	359
1221	sp P97283 DPOD1_MESAU	AKVQSYEREEDLLQAWPNFILAMPDVIITGYNIQNFDPYIISRAQTLKVDRFPFLGRVT	421
1222	sp O48901 DPOD1_SOYBN	VDVMPFETEREVLLAWRDILRVPDVIITGYNICKFDLPYIIERALNLKIAEFPIILGRIR	386
1223		: : : * * :: ***: **::**::* . * : . : **	
1224			
1225			
1226	sp P30316 DPOD_SCHPO	NFFSVAKETTFSSKAYGTRESKTTISIPGRLQLDMLQVMQRDFKLRSYSLNAVCSQFLGEQ	469
1227	sp P15436 DPOD_YEAST	TVKQEIKEVSFSSKAYGTRETKNVNIDGRLQLDLLQFIQREYKLRSYTLNAVSAHFLGEQ	490
1228	sp P28339 DPOD1_BOVIN	GLRSNIRESSFSQSRQTGRRDSKVSMVGRVQMDMLQVLLREYKLRSYTLNAVVSFHLGEQ	484
1229	sp P28340 DPOD1_HUMAN	GLCSNIRDSSFQSKQTGRRDTKVSMVGRVQMDMLQVLLREYKLRSYTLNAVVSFHLGEQ	485
1230	sp O54747 DPOD1_RAT	GLRSNIRDSSFQSRQVGRRDSKVSMVGRVQMDMLQVLLREYKLRSYTLNAVVSFHLGEQ	481
1231	sp P30315 DPOD1_PLAFK	NVASTVKDSFSSKQFGTHTKEYNIFGRIQFDVVDLIKRDYKLRSYTLNAVVSFEFLKEQ	479
1232	sp P54358 DPOD1_DROME	NIRSVIKEQMLQSKQGRRENKYVNFEGRVFDDLFLVLLRDYKLRSYTLNAVVSYHFLGEQ	469
1233	sp P52431 DPOD1_MOUSE	GLRSNIRDSSFQSRQVGRRDSKVISMVGRVQMDMLQVLLREHKLRSYTLNAVVSFHLGEQ	483
1234	sp P90829 DPOD1_CAEEL	EKGSVVRDAAISSKQMSRVNKSIDIHGRIIFDVLQVVLRDYKLRSYTLNVSYSYQFLSEQ	455
1235	sp Q9LVN7 DPOD1_ARATH	NSRVVRDSTFSSRQQGIRESKETTIEGRFQFDLIQAIHRDHKLSYSLNVSVAHFLSEQ	479
1236	sp Q9LRE6 DPOD1_ORYSJ	NSRVVRDSTFSSRQYGMRESKDVAVEGRVQFDLLQAMQRDYKLSYSLNVSVAHFLSEQ	489
1237	sp P46588 DPOD_CANAL	RVKQEIKDAVFSSRAYGTRENKVVNIDGRMQLDLLQFIQREYKLRSYTLNVSVAHFLGEQ	419
1238	sp P97283 DPOD1_MESAU	GLRSNIRDSSFQSRQVGRRDSKVSMVGRVQMDMLQVLLREHKLRSYTLNAVVSFHLGEQ	481
1239	sp O48901 DPOD1_SOYBN	NSRVVRDSTFSSRQYGTRESKEVAVEGRVTFDDLQVMQRDYKLSYSLNVSSSHFLSEQ	446
1240		:: :*: * : : . **::**::* : *::** **::** * . . ** *	
1241			

242
243 sp|P30316|DPOD_SCHPO KEDVHYSIIIDLQNGTADSRRLAVYCLKDAYLPQRLMDKLMCFVNYTEMARVTVGVPI 529
244 sp|P15436|DPOD_YEAST KEDVHYSIIIDLQNGDSETRRLAVYCLKDAYLPRLRMEKLMALVNYTEMARVTVGVPI 550
245 sp|P28339|DPOD1_BOVIN KEDVQHSIIIDLQNGNDQTRRLAVYCLKDAFLPLRLLERLMVLVNAMEMARVTVGVPI 544
246 sp|P28340|DPOD1_HUMAN KEDVQHSIIIDLQNGNDQTRRLAVYCLKDAFLPLRLLERLMVLVNAMEMARVTVGVPI 545
247 sp|O54747|DPOD1_RAT KEDVQHSIIIDLQNGNEQTRRLAVYCLKDAFLPLRLLERLMVLVNNVEMARVTVGVPI 541
248 sp|P30315|DPOD1_PLAFK KEDVHYSIMNDLQNESPESEKRIATYCIKDGVLPLRLDKLLFIYNYVEMARVTVGVPI 539
249 sp|P54358|DPOD1_DROME KEDVHHSIIIDLQNGDEQTRRLAMVCLKDAYLPRLRLEKLMALVNYMEMARVTVGVPI 529
250 sp|P28341|DPOD1_MOUSE KEDVQHSIIIDLQNGNDQTRRLAVYCLKDAFLPLRLLERLMVLVNNVEMARVTVGVPI 543
251 sp|P90829|DPOD1_CAEEL KEDVEHNIIDLQNGDEQTRRLAQYCLKDAYLPRLRLDKLMSIINYIEMARVTVGVPI 515
252 sp|Q9LVN7|DPOD1_ARATH KEDVHHSIIIDLQNGNAETRRRLAVYCLKDAYLPQRLDKLMFIYNYVEMARVTVGVPI 539
253 sp|Q9LRE6|DPOD1_ORYSJ KEDVHHSIIIDLQNGNSETRRLAVYCLKDAYLPQRLDKLMSIINYIEMARVTVGVPI 540
254 sp|P46588|DPOD_CANAL KEDVQHSIIIDLQNGTKETRRRLAVYCLKDAFLPLRLDKLMLVNYTEMARVTVGVPI 479
255 sp|P97283|DPOD1_MESAU KEDVQHSIIIDLQNGNEQTRRLAVYCLRDAFLPLRLLERLMVLVNNVEMARVTVGVPI 541
256 sp|O48901|DPOD1_SOYBN KEDVHHSIIIDLQNGNAETRRRLAVYCLKDAYLPQRLDKLMFIYNYVEMARVTVGVPI 506
257 ****.:*: *** :*:*** *:*: ** *::*: : * *****.*
258
259
260
261 sp|P30316|DPOD_SCHPO LLARGQIKIVISQLFRKALQHDLVVPNIRVNGT-DEQYEGATVIEPIKGYDPIATLDF 588
262 sp|P15436|DPOD_YEAST LLARGQIKVVSQFLRKCLIEDTVIPNMQSQAS-DDQYEGATVIEPIRGYDVPPIATLDF 609
263 sp|P28339|DPOD1_BOVIN LLSRGQQVKVVSQFLRQAMRQGLMPVVKTEGG--EDYT GATVIEPLKGYDVPPIATLDF 602
264 sp|P28340|DPOD1_HUMAN LLSRGQQVKVVSQFLRQAMHEGLMPVVKSEGG--EDYT GATVIEPLKGYDVPPIATLDF 603
265 sp|O54747|DPOD1_RAT LLSRGQQVKVVSQFLRQAMREGLMPVVKTEGG--EDYT GATVIEPLKGYDVPPIATLDF 599
266 sp|P30315|DPOD1_PLAFK LLTRGQQIKVTSQLYRKCKELNYIPSTYMKVNTNEKYEGATVIEPIKGYDPIATLDF 599
267 sp|P54358|DPOD1_DROME LLTRGQQIKVLSQFLRKAKTKGFIMPSYTSQGS-DEQYEGATVIEPKRGYADPISTLDF 588
268 sp|P52431|DPOD1_MOUSE LLTRGQQVKVVSQFLRQAMRQGLMPVVKTEGG--EDYT GATVIEPLKGYDVPPIATLDF 601
269 sp|P90829|DPOD1_CAEEL LLTKGQQIKILSMMLRRCKQNNFLPVIANSQDGEYEGATVIDPIRGFYNEPIATLDF 575
270 sp|Q9LVN7|DPOD1_ARATH LLARGQSIKVLSQLLRKGQKNLVLPAKQSGSEQGTYE GATVLEARTGFYEKPIATLDF 599
271 sp|Q9LRE6|DPOD1_ORYSJ LLSRGQSIKVLSQLLRKAKQKNLVIPNIKQASQDQTFEGATVLEARAGFYEKPIATLDF 609
272 sp|P46588|DPOD_CANAL LLSRGQQIKVVSQFLRKCLQEDIVPNLKSEGS-NEEYEGATVIEPERGYDVPPIATLDF 538
273 sp|P97283|DPOD1_MESAU LLSRGQQVKVVSQFLRQAMRQGLMPVVKTEGG--EDYT GATVIEPLKGYDVPPIATLDF 599
274 sp|O48901|DPOD1_SOYBN LLSRGQSIKVLSQLRRARQKNLVIPNAKQAGSEQGTFE GATVLEARAGFYEKPIATLDF 566
275 ***:*** :*: * :* : . :* : *****.: ** ***
276
277
278 sp|P30316|DPOD_SCHPO SSLYPSIMQAHNLCYTTLDSNTAEELL---KLKQDV-DYSVTPNGDYFVKPHVRKGGLP 643
279 sp|P15436|DPOD_YEAST NSLYPSIMMAHNLCYTTLCKNKATVERL---NLKIDE-DYVITPNGDYFVTKRRRGILP 664
280 sp|P28339|DPOD1_BOVIN SSLYPSIMMAHNLCYTTLRLPGAQKL---GLTED--QFIKPTGDEFVKASVRKGGLP 656
281 sp|P28340|DPOD1_HUMAN SSLYPSIMMAHNLCYTTLRLPGAQKL---GLTED--QFIKPTGDEFVKASVRKGGLP 657
282 sp|O54747|DPOD1_RAT SSLYPSIMMAHNLCYTTLRLPGAQKL---GLKPD--EFIKPTGDEFVKASVRKGGLP 653
283 sp|P30315|DPOD1_PLAFK ASLYPSIMIAHNLCYTTLKSNH-EVS---DLQNDITTIQKGNLKFVKNNVKGILP 654
284 sp|P54358|DPOD1_DROME ASLYPSIMMAHNLCYTTLVLGRTREKLQQENLQDD--QVERTPANNYFVKSEVRRGGLP 646
285 sp|P52431|DPOD1_MOUSE SSLYPSIMMAHNLCYTTLRLPGAQKL---GLKPD--EFIKPTGDEFVKASVRKGGLP 655
286 sp|P90829|DPOD1_CAEEL ASLYPSIMIAHNLCYTTLKSP---Q---GVENE--DYIRTPSGQYFATKSKRRGGLP 625
287 sp|Q9LVN7|DPOD1_ARATH ASLYPSIMMAYNLCYTTLVPEDVRKL---NLPE--HVTKTPSGETFVKQTLQKGILP 653
288 sp|Q9LRE6|DPOD1_ORYSJ ASLYPSIMMAYNLCYTTLVPEDARKL---NLPE--SVNKTTPSGETFVKPDVQKGILP 663
289 sp|P46588|DPOD_CANAL SSLYPSIMMAHNLCYTTLNKNKSIKAF---GLTED--DYTKTPNGDYFVHNSLRKGILP 592
290 sp|P97283|DPOD1_MESAU SSLYPSIMMAHNLCYTTLRLPGAQKL---GLKPD--EFIKPTGDEFVKSSVRKGGLP 653
291 sp|O48901|DPOD1_SOYBN ASLYPSIMMAYNLCYTTLVPEDARKL---NLPE--SVNRTTPSGETFVKSNLQKGILP 620
292 *****.*:***** * : : . * . :*:***
293
294
295
296 sp|P30316|DPOD_SCHPO IILADLLNARKKAKADLKKETDPFKKAVLDGRQLALKVSANSVYGFTGATN-GRLPCLIAI 702
297 sp|P15436|DPOD_YEAST IILDELISARKRAKDLRDEKDPFKRDVLDGRQLALKKISANSVYGFTGATV-GLKPCIAI 723
298 sp|P28339|DPOD1_BOVIN QILENLLSARKRAKAEALAKETDPLRRQVLDGRQLALKVSANSVYGFTGAQV-GRLPCLIEI 715
299 sp|P28340|DPOD1_HUMAN QILENLLSARKRAKAEALAKETDPLRRQVLDGRQLALKVSANSVYGFTGAQV-GLKPCLEI 716
300 sp|O54747|DPOD1_RAT QILENLLSARKRAKAEALAKETDPLRRQVLDGRQLALKVSPNSVYGFTGATV-GLKPCLEI 712
301 sp|P30315|DPOD1_PLAFK LIVEELIEARKKVKLLIKNEKNNTKMLVLDGRQLALKKISANSVYGYTGASSGQGLPCIEV 714
302 sp|P54358|DPOD1_DROME EILESLLAARKRAKNDLVETDPFKRKVLVDGRQLALKKISANSVYGFTGAQV-GLKPCLEI 705
303 sp|P52431|DPOD1_MOUSE QILENLLSARKRAKAEALAKETDPLRRQVLDGRQLALKVSANSVYGFTGAQV-GLKPCLEI 714
304 sp|P90829|DPOD1_CAEEL EILEDILAAARKRAKNDMKNEKDEFKRMVYLDGRQLALKKISANSVYGYTGATV-GLKPCLEI 684
305 sp|Q9LVN7|DPOD1_ARATH EILEELLTARKRAKADLKEAKDPLEKAVLDGRQLALKKISANSVYGYTGATV-GLKPCLEI 712
306 sp|Q9LRE6|DPOD1_ORYSJ EILEELLTARKRAKADLKEAKDPFERAVLDGRQLALKKISANSVYGYTGATV-GLKPCLEI 722
307 sp|P46588|DPOD_CANAL TILDELLTARKKAKADLKKETDPFKKDVLDGRQLALKKISANSVYGYTGATV-GLKPCIAI 651
308 sp|P97283|DPOD1_MESAU QILENLLSARKRAKAEALAKETDPLRRQVLDGRQLALKVSANSVYGFTGAQV-GLKPCLEI 712
309 sp|O48901|DPOD1_SOYBN EILEELLTARKRAKADLKEAKDPLEKAVLDGRQLALKKISANSVYGYTGATI-GLKPCLEI 679
310 *: .: *****.* : : : * *****.* *****.* *:***** :
311
312

Fig. 2J CLUSTAL format for T-COFFEE Version of Delta Polymerases (Eukaryotes)
The possible catalytic region is underlined and shaded yellow. Only regions showing highly conserved blocks are shown.

[illegible]

1453			
1454	gi 285814403 tpg DAA10297.1	YAFEKPHLSEYVSVDATYYLYMKYVHPFIFSLCTIIPLNPDETLRKG	511
1455	gi 118841 sp P21951.1 DPOE_YEA	YAFEKPHLSEYVSVDATYYLYMKYVHPFIFSLCTIIPLNPDETLRKG	511
1456	gi 259485986 tpe CBF83469.1	YASERPQTAEYVSVDATYYLYMKYIHPFIFSLCTILPLNPDDTLRKG	480
1457	gi 194214443 ref XP_001915746.	MATEQPQTATYVSVDATYYLYMKYVHPFIFALCTIIPMEPDEVLRKG	530
1458	Q07864-1	MATEQPQTATYVSVDATYYLYMKYVHPFIFALCTIIPMEPDEVLRKG	496
1459		* *: * *: *****:*****:*****:..* *: *	
1460			
1461			
1462			
1463	gi 285814403 tpg DAA10297.1	TGTLCEMLLMVQAYQHNIILLPNKHHTDPIERFYDG-HLLESETYVGGHVES	560
1464	gi 118841 sp P21951.1 DPOE_YEA	TGTLCEMLLMVQAYQHNIILLPNKHHTDPIERFYDG-HLLESETYVGGHVES	560
1465	gi 259485986 tpe CBF83469.1	TGTLCEMLLMVQAYKGNIVLPNKHKDPPEAFYEG-HLLESETYVGGHVES	529
1466	gi 194214443 ref XP_001915746.	SGTLCEALLMVQAFHANIIFPNKQEQEFNKLTDGHHVLDSETYVGGHVEA	580
1467	Q07864-1	SGTLCEALLMVQAFHANIIFPNKQEQEFNKLTDGHHVLDSETYVGGHVEA	546
1468		.*****:*****: * *: * *: : : : : . * *:*****:	
1469			
1470			
1471			
1472	gi 285814403 tpg DAA10297.1	LEAGVFRSDLKNEFKIDPSAIDELLQELPEALKFSVEVENKSSVDKVTNF	610
1473	gi 118841 sp P21951.1 DPOE_YEA	LEAGVFRSDLKNEFKIDPSAIDELLQELPEALKFSVEVENKSSVDKVTNF	610
1474	gi 259485986 tpe CBF83469.1	IEAGVFRSDIPVPFNIDPTAVDELLRDLDAALKFSIEVEEKKSLDDVTNY	579
1475	gi 194214443 ref XP_001915746.	LESVFRSDIPCRFRMNPAADFLLQRVEKTLRHAIEEEEKVPVEQVTNF	630
1476	Q07864-1	LESVFRSDIPCRFRMNPAADFLLQRVEKTLRHAIEEEEKVPVEQVTNF	596
1477		* *:*****: * *: * *: * *: : : * *: * *: * *: * *: *	
1478			
1479			
1480	gi 285814403 tpg DAA10297.1	EEIKNQITQKLELLEKNNIRNELPLIYHVDVASMYPNIMTTNRLQPSIK	660
1481	gi 118841 sp P21951.1 DPOE_YEA	EEIKNQITQKLELLEKNNIRNELPLIYHVDVASMYPNIMTTNRLQPSIK	660
1482	gi 259485986 tpe CBF83469.1	EEVKGQIAKLLTDLRENPHRNEVPLIYHLDVASMYPNIMTTNRLQPSLI	629
1483	gi 194214443 ref XP_001915746.	QEVCDQIKTKLTSKLDVFNRIECPLIYHLDVGAMYPNIIILTNRLQPSAMV	680
1484	Q07864-1	EEVCDQIKTKLTSKLDVFNRIECPLIYHLDVGAMYPNIIILTNRLQPSAMV	646
1485		* *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: *	
1486			
1487			
1488	gi 285814403 tpg DAA10297.1	AERDCASCDFNRPCKTCARKLWAWRGEFFPSKMDEYNMIKRALQNETFP	710
1489	gi 118841 sp P21951.1 DPOE_YEA	AERDCASCDFNRPCKTCARKLWAWRGEFFPSKMDEYNMIKRALQNETFP	710
1490	gi 259485986 tpe CBF83469.1	QESNCAACDFNRPCKTCDRRLPWAWRGEFFPAKRDEYNMIRQAVQNERFP	679
1491	gi 194214443 ref XP_001915746.	DEATCAACDFNKPANCQRKMAWQWRGEFMPASRSEYHRIHQHLESEKFP	730
1492	Q07864-1	DEATCAACDFNKPANCQRKMAWQWRGEFMPASRSEYHRIHQHLESEKFP	696
1493		* *:*****: * *: * *: * *: * *: * *: * *: * *: * *: * *: *	
1494			
1495			
1496	gi 285814403 tpg DAA10297.1	NKNKFSKKKVLTFDELSYADQVIHIKKRLTEYSRKVYHRVKVSEIVEREA	760
1497	gi 118841 sp P21951.1 DPOE_YEA	NKNKFSKKKVLTFDELSYADQVIHIKKRLTEYSRKVYHRVKVSEIVEREA	760
1498	gi 259485986 tpe CBF83469.1	GRTKKSP--MRAFTELSAAEQAAIVKKRLQDYSKKIYHKIHDSTMTVREA	727
1499	gi 194214443 ref XP_001915746.	PLTPEGP--ARAFHELSSREEQAKYEKRLADYCRKAYKKIHVTKVEERLT	778
1500	Q07864-1	PLFPEGP--ARAFHELSSREEQAKYEKRLADYCRKAYKKIHVTKVEERLT	744
1501		. : * * * : * . * : * : * : * : * : * : * : *	
1502			
1503			
1504	gi 285814403 tpg DAA10297.1	IVCQRENPFYVDTVKSFRDRRYEFKGLAKTWKGNLSKIDPSDK--HARDE	808
1505	gi 118841 sp P21951.1 DPOE_YEA	IVCQRENPFYVDTVKSFRDRRYEFKGLAKTWKGNLSKIDPSDK--HARDE	808
1506	gi 259485986 tpe CBF83469.1	IICQRENPFYVDTVRSFRDRRYDFKGLQKQVWKGKTESLKSSGAPAAEIEE	777
1507	gi 194214443 ref XP_001915746.	TICQRENSFYVDTVRAFRRDRRYEFKGLHKVWKKKLSAAVEVGD-AAEVKR	827
1508	Q07864-1	TICQRENSFYVDTVRAFRRDRRYEFKGLHKVWKKKLSAAVEVGD-AAEVKR	793
1509		.*****:*****:*****: * *: * *: * *: * *: * *: * *: *	
1510			
1511			
1512	gi 285814403 tpg DAA10297.1	AKKMIIVLYDSLQLAHKVIILNSFYGYVMRKGSRWYSMEMAGITCLTGATII	858
1513	gi 118841 sp P21951.1 DPOE_YEA	AKKMIIVLYDSLQLAHKVIILNSFYGYVMRKGSRWYSMEMAGITCLTGATII	858
1514	gi 259485986 tpe CBF83469.1	AKKMIIVLYDSLQLAHKVIILNSFYGYVMRKGSRWYSMEMAGITCLTGATII	827
1515	gi 194214443 ref XP_001915746.	CRNMEIVLYDSLQLAHKCIILNSFYGYVMRKGSRWYSMEMAGIVCFTGANII	877
1516	Q07864-1	CKNMEIVLYDSLQLAHKCIILNSFYGYVMRKGSRWYSMEMAGIVCFTGANII	843
1517		..* *****:*****:*****: * *: * *: * *: * *: * *: * *: *	
1518			
1519			
1520			
1521			
1522			
1523			

1524	gi 285814403 tpg DAA10297.1	QMARALVERVGRPLELDTDGIWCILPKSFPEYFFTTLENGKK--LYLSYP	906
1525	gi 118841 sp P21951.1 DPOE_YEA	QMARALVERVGRPLELDTDGIWCILPKSFPEYFFTTLENGKK--LYLSYP	906
1526	gi 259485986 tpe CBF83469.1	QMARLVERIGRPLELDTDGIWCMLPGTFPENFSFTLKNGKK--LGISYP	875
1527	gi 194214443 ref XP_001915746.	QARELIEQIGRPLELDTDGIWCVLPNSFPENFVVKTSVKKPKVTISYP	927
1528	Q07864-1	QARELIEQIGRPLELDTDGIWCVLPNSFPENFVVKTTNVKKPKVTISYP	893
1529		*** *:*****:*** :*** : .. * * : :***	
1530			
1531			
1532			
1533	gi 285814403 tpg DAA10297.1	CSMLNVRVHQKFTNFOYQELKDPLNYIYETHSENTIFFEVDGPYKAMILP	956
1534	gi 118841 sp P21951.1 DPOE_YEA	CSMLNVRVHQKFTNFOYQELKDPLNYIYETHSENTIFFEVDGPYKAMILP	956
1535	gi 259485986 tpe CBF83469.1	CVMLNHLVHGSYTNFOYQSLANPAFTFYETHSENSIFFEVDGPYRAMILP	925
1536	gi 194214443 ref XP_001915746.	GAMLNIMVKEGFTNDYOYQLAEPSSLTIVTRSENSIFFEVDGPYLAMILP	977
1537	Q07864-1	GAMLNIMVKEGFTNDYOYQLAEPSSLTIVTRSENSIFFEVDGPYLAMILP	943
1538		*** *:***:*** * * * . * *:***:***** *****	
1539			
1540			
1541	gi 285814403 tpg DAA10297.1	SSKEEGKGIKKRYAVFNEDGSLAELKGFEVKRRGELQLIKNFQSDIFKVF	1006
1542	gi 118841 sp P21951.1 DPOE_YEA	SSKEEGKGIKKRYAVFNEDGSLAELKGFEVKRRGELQLIKNFQSDIFKVF	1006
1543	gi 259485986 tpe CBF83469.1	TSKEEDKNLKKRYAVFNDDGSLAELKGFEVKRRGELQLIKIFQTIQIFKF	975
1544	gi 194214443 ref XP_001915746.	ASKEEGKKLKKRYAVFNEDGSLAELKGFEVKRRGELQLIKIFQSVFEAF	1027
1545	Q07864-1	ASKEEGKKLKKRYAVFNEDGSLAELKGFEVKRRGELQLIKIFQSVFEAF	993
1546		:****.* :*****:*****:*****:*** *:..*: *	
1547			
1548			
1549	gi 285814403 tpg DAA10297.1	LEGDTLEGCSAVASVCNRWLDVLDHGLMLEDEDLVSLICENRSMSTKL	1056
1550	gi 118841 sp P21951.1 DPOE_YEA	LEGDTLEGCSAVASVCNRWLDVLDHGLMLEDEDLVSLICENRSMSTKL	1056
1551	gi 259485986 tpe CBF83469.1	LEGTTLAETYSAAVARWLDVLYSHGATLADEELIELISENRSMTKL	1025
1552	gi 194214443 ref XP_001915746.	LKGSTLEEVYGSVAKVADYWLVDVLYSKAANMPDSELFELISENRSMSRKL	1077
1553	Q07864-1	LKGSTLEEVYGSVAKVADYWLVDVLYSKAANMPDSELFELISENRSMSRKL	1043
1554		*:* ** *:*** *: ***** .. :* :..**.******:..*	
1555			
1556			
1557	gi 285814403 tpg DAA10297.1	KEYEGQKSTSITTARRLGDFLGEDMVKDKGLOCKYIISSKPFNAPVTERA	1106
1558	gi 118841 sp P21951.1 DPOE_YEA	KEYEGQKSTSITTARRLGDFLGEDMVKDKGLOCKYIISSKPFNAPVTERA	1106
1559	gi 259485986 tpe CBF83469.1	EEYGNQKSTSITTARRLAEFLGEDMVKDGLNCKYIIISARPRNTPVTERA	1075
1560	gi 194214443 ref XP_001915746.	EDYGEQKSTSISTAKRLAEFLGDQMVKDAGLSCRYIISRKPEGSPVTERA	1127
1561	Q07864-1	EDYGEQKSTSISTAKRLAEFLGDQMVKDAGLSCRYIISRKPEGSPVTERA	1093
1562		::* *****:***:***:***:*** ** *:***** :* ..*****	
1563			
1564			
1565	gi 285814403 tpg DAA10297.1	IPVAIFSADIPIKRSFLRRWTLDPSSLEDLDIRTIIDWGYRERLGSIAIQK	1156
1566	gi 118841 sp P21951.1 DPOE_YEA	IPVAIFSADIPIKRSFLRRWTLDPSSLEDLDIRTIIDWGYRERLGSIAIQK	1156
1567	gi 259485986 tpe CBF83469.1	IPVTIFSADSIKRRHFLRKWLKD-DPGMDPRSDVIDWLYLERLGSVQK	1124
1568	gi 194214443 ref XP_001915746.	IPLAIFQAEPTVRKHFLRKWLKSSSLQFDIRTILDWYIERLGSIAIQK	1177
1569	Q07864-1	IPLAIFQAEPTVRKHFLRKWLKSSSLQFDIRAILDWYIERLGSIAIQK	1143
1570		***:***.* :..:***:* . . :*:***:*** ** *****:***	
1571			
1572			
1573	gi 285814403 tpg DAA10297.1	IITIPAALQCVSNPVPRVEHPDWLKRKIATKEDKFKQTSLTKFSSK--TK	1204
1574	gi 118841 sp P21951.1 DPOE_YEA	IITIPAALQCVSNPVPRVEHPDWLKRKIATKEDKFKQTSLTKFSSK--TK	1204
1575	gi 259485986 tpe CBF83469.1	LITIPAALQKIRNPVPRVAHPWLRRIKQDDRFKQVKMTDMFGK--SE	1172
1576	gi 194214443 ref XP_001915746.	IITIPAALQCVKNPVPRVKHPDWLHKKLLEKNDIYKQKKISELFILEGKR	1227
1577	Q07864-1	IITIPAALQCVKNPVPRVKHPDWLHKKLLEKNDVYKQKKISELFTLEGRR	1193
1578		***** : ***** ***:***:***:***:***:***:***	
1579			
1580			
1581	gi 285814403 tpg DAA10297.1	VLEVFVTTINGKVNITFHIPTKIYMKFKSQTMPLQIKINCLIEKSSASLP	1384
1582	gi 118841 sp P21951.1 DPOE_YEA	VLEVFVTTINGKVNITFHIPTKIYMKFKSQTMPLQIKINCLIEKSSASLP	1384
1583	gi 259485986 tpe CBF83469.1	IVRAFLVLDIRKIHALTIKVPRCVYINLKQDSLDPVEVPECEVEKVNHTLP	1359
1584	gi 194214443 ref XP_001915746.	LFRLWAVISSDLICYIKLNI PRVFYVNRQ---VAKAEEGSPYRKVNRLP	1423
1585	Q07864-1	LFRLWALVGSDDLHCIRLSIPRVFYVNRQ---VAKAEEGASRYKVNRLP	1389
1586		 : : * : * : : . * . **	
1587			
1588			



The possible catalytic region is underlined and shaded yellow.

Only regions showing highly conserved blocks are shown.

3.3 Conserved amino acid(s) in DNA polymerases

Multiple sequence alignments of known prokaryotic and eukaryotic polymerases have clearly indicated two conserved sets of amino acids, viz., one consisting of the proposed catalytic site amino acid K and the other is an YG pair. In some cases the K is replaced by the equivalent amino acid R and in the YG pair, the G is replaced by A/I (Table 1). It is interesting to note that a distance conservation is also observed in all these polymerases, i.e., the YG pair is positioned around 6-9 amino acids downstream from the catalytic amino acid K (Table 1).

Interestingly in all eukaryotic polymerases the second amino acid from the catalytic K is mostly a branched chain amino acid such as V/I/L. Similarly pol IV and pol V also use a branched chain amino acid. The steric gate amino acid Y is strictly followed by G in all cases except eukaryotic beta polymerases. In polymerase II, instead of a K, an R is conserved. Unlike the pol I the pol II cannot use a nicked duplex templates of DNA; it usually fills short gaps in DNA which act as a block to DNA polymerase and thus helps reinitiate DNA synthesis in UV induced cells. In some of the prokaryotic polymerases like pol IV and pol V, the steric gate amino acid Y is not found at the expected distance from the possible catalytic amino acid K, which probably explains the error-prone nature of the polymerases (Table 1).

Table 1. Conserved regions around the catalytic and steric-gate amino acids

(Numbering from the catalytic amino acid K/R).

Prokaryotic polymerases		Eukaryotic polymerases	
pol I	A _V /K ¹ A _{T/S} V/I/MN/SY/FGLI/A ⁹ Y ⁸ G	Alpha	L/I/K ¹ LTANSMY ⁸ GCLG
pol II	K/R/M/R ¹ H/L/L/VXPXF/Y ⁷ G	Beta	K/S/K ¹ V/I/LD/F/S/P/E/K ⁶ I/LA/GT
pol III	NK/R ¹ S/AHS/A/A/Y ⁷ G/A _s	Gamma	AK ¹ I/VFNYG/A _R IY ⁹ GAG
pol IV	L/M/AK ¹ L/M/A/GSXXXK ⁹ XG ¹¹	Delta	LK ¹ I/VSANSVY ⁸ GFTGA
pol V	A/s/K ¹ L/VAN/M/QXXXXX	Epsilon	HK ¹ V/cILNSFY ⁸ GYVMRKG
Viral	A _N /K ¹ /RXF/IXXGXL ⁹ Y ⁸ GA/s		

Less common amino acids around the catalytic amino acid K, are shown in lower font sizes. X represents any amino acid that is not conserved.

3.4 Absence of YG pair in error prone polymerases

Another interesting observation is that the error prone *E. coli* polymerases pol IV and V did not show any YG pair at the expected positions from the catalytic K (Table 1). These polymerases were reported in 1999 and are grouped under error prone DNA polymerases [3]. (DNA polymerase IV is induced in stationary phase cells and is involved in adaptive mutagenesis whereas the DNA polymerase V replicates past gaps in DNA. It is made of two different subunits (*umuC* and *umuD'*) and is involved in SOS DNA repair pathway, when *E. coli* is exposed to high levels of radiation or mutagens and a major damage is done to the DNA).

Thus, polymerase IV (*dinB*) and Pol V (*umuD'C*) belong to the newly discovered class of Y DNA polymerases [3]. Polymerases that belong to Y family are often referred to as specialized or error-prone DNA polymerases to distinguish them from the previously described DNA polymerases (pol I, pol II, and pol III) that are essentially involved in DNA replication or error-free DNA repair. These Y-family DNA polymerases are characterized by their ability to replicate DNA, through chemically damaged template bases, or to elongate mismatched primer termini. These properties stem from their capacity to accommodate and use distorted primer templates within their active site and from the lack of an associated exonuclease activity. It is Interesting to note

that the explicit absence of the YG steric gate pair in these polymerases explains their error prone nature.

In the eukaryotic polymerase β , a nuclear DNA repair polymerase that plays a key role in base-excision repair, the G in YG pair is replaced with a branched chain amino acid, like I or L. The implication of such a replacement is not known.

3.5 Distance conservation between catalytic K and YG pair

It is interesting to note that the catalytic amino acid K and the gate keeper pair YG are completely conserved in different polymerases from diverse group of organisms. The mismatched regions in some of the polymerases were aligned as suggested by Palanivelu [4]. Table 2 further corroborates the above findings. 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. It is interesting to note that irrespective of the type of polymerases and their origin, all 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.

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

T4 DNA pol	546 ATLANTNQLNR ^{K1} ILINSL ^{Y8} GALGNIH
Human HSV 1	800 AVLLDKQQA ^{K1} VVCNSV ^{Y8} GFTGVQH
<i>E. coli</i> DNA pol I	748 TVTSEQRRSA ^{K1} AINFGLI ^{Y9} GMSAFGLAR
<i>E. coli</i> DNA pol II	482 RQGNKPLSQAL ^{K1} IIMNAF ^{Y8} GVLGTTA
<i>E. coli</i> DNA pol III (alpha subunit)	663 YPDVQWQHESL ^{K1} PVLEPT ^{Y8} GIILYQE
<i>P. furiosus</i> DNA pol	477 KILLDYRQKA ^{K1} LLANSF ^{Y8} GYYGYAK
Yeast Alpha DNA pol	933 RVQC DIRQQAL ^{K1} LTANS ^{MY8} GCLGYVN
Human Alpha DNA pol	939 ILQYDIRQKAL ^{K1} LTANS ^{MY8} GCLGFSY
Human Gamma DNA pol	917 TTVGISREHA ^{K1} IFNYGRI ^{Y9} GAGQPFAER
Human Delta DNA pol (Catalytic subunit)	683 RQVLDGRQLAL ^{K1} VSANSV ^{Y8} GFTGAQV
Human Epsilon DNA pol	798 EVLYDSLQLAH ^{K1} CILNSF ^{Y8} GYVMRKGAR
<i>A. thaliana</i> Delta DNA pol (Catalytic subunit)	679 KAVLDGRQLAL ^{K1} ISANSV ^{Y8} 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 [4].

3.6 Other conserved motif(s) in prokaryotic and eukaryotic polymerases

Apart from these highly conserved regions, not much conservation is observed among different polymerase families. However, there are conserved motifs among the members of the same family. Prokaryotic polymerase pol V and viral polymerases did not show much



conservation among themselves. Eukaryotic polymerases δ and ϵ showed maximum number of conserved motifs. ϵ polymerases showed long stretches of conserved motifs among them (Table 3).



Notable one is the DXD motif which is found in all eukaryotic polymerases but only in one of the prokaryotic polymerases viz., in the replicative polymerase, pol III. This triad is also implicated in catalysis. Delta and Epsilon polymerases have two such triads.

Table 3. Other highly conserved regions found among different prokaryotic and eukaryotic polymerases.

Table 3A. Prokaryotic polymerases	
pol I	EADDV/I IXS/T SP/TKQL TGRLLS PNLQNIP DYS QIEL/MRL/I/VL/MAH IQGS/T QVHDEL VPL
pol II	V/I/V/IG V/I/LED PGNHD PQP GHI/VH V/I/NS
pol III	GPGRGS LL/IFERFLNPER PDX DV/ID F LIKXFDLGL RPPG IV/L/IYQEQV/I
pol IV	I/VIHV/ID FFA RGV/IV/I LSL/IDEA/G GV/IG VRLL/VGV/L/I
pol V	D/EV/IWGV/IGR/K
Viral	YNV/I V/ILN ESI
Table 3B. Eukaryotic polymerases	
Alpha	LFGK LLHXF PDK GGLVLD/EP ILLDFNSLYPSIIQEY/FNICFTTV/I SRFY VIYG DTD SIMI KKKYAA KGLD RRDW/Y/F
Beta	CGSY/F DXD VL KIL RRXDIR GVLY
Gamma	KLPH GTITRAVEXTWLTASN RLGSELK F/LVGADVDSQ/EELWI GT/SD TKG SRV/IN VDY LHL HDEV/IRYL AFFSX DID LRKE TPSN
Delta	LSFDIEC FPE DPVIQI DPDI/V GYN FDLPYL SYS/TLN KEDVDLQN YCLKD EMARVTGXP RGQ GATVL/I TLDF SLYPSIMMAXNLCYXTL GL/ILP ARKR GXLPLC VI/VYG DTD SVM PIK/RLEFEKVYXPYLLIXKKRYAGL KGI/L LVITK V/LGDRVXY EDP
Epsilon	GWLTMN VDYYFIQD/ED/EG PYFXIA KEDL NHLL/V DXREYDVPYHIRLSID DPVVL/MAFDIETTKXPLKFPDXXDQIMMI SYMIDGXY/FLITNREIVSEIDFEF/YTPKPEY I/VFNEF/WFEH TYNGDFFDWPFV/I EIGF HMDCCXXWXRDSYLPXGSXXXLKA KLGYPXELDPE YSVSDAVATYYLYMKYVHPFIFALCTIIPMEPDEXLRKGS/TGTLCEXLLMVQAY/F NII/L/VXPKN HVLDSEYVGGHVE GVFRSDI/L VTNF/Y IYHLDVG/AXMPNIXXNRLQP CAXCDFNKPX WXWRGEFXP FXELS V/ICQRENXYFVDTVR/KXFRDRRYD/EFKG ARXLIEXIGRPLEL DTD GIWCXLPXSFP L/ISYP MLN QYQYXTR/HSENS/TIFFEVDGPYXAMILPXSKEE KKRYAVFNEDGSLAELKGFEVKRRGELXLIKXFS/T WLDVL LIXENRSMR/KXL QKSTSIS/TTAK/RRLXEFGLDXMVKDXGLXCR/KYIIS PVTERA IPLXIF FLRKW YYXERLGSXIQTIPAAALQXXXPVPRVXKHPDWL VNTILXS PVLPG NPXLEXI/VK DL/ID QDLXCK/RC

NB: X represents any amino acid Smaller font sizes represent less frequently used amino acids.

3.7 Three critical pairs in different types of polymerases and their possible functions

These analyses have revealed three critical pairs of amino acids in all these polymerases belonging to different types and origins. Table 4 shows these three critical pairs from different representative polymerases.

Table 4. Critical pairs of amino acids found in different types of polymerases

Enzyme	Lys//Arg	Tyr/Phe-Gly	Asp//Asp
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

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 (Fig.2C)

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 completely conserved P known to act as a helix breaker may introduce the necessary sharp kinks into polypeptide backbone (Fig. 2D).

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.

- The pair 1 involves in polymerization, (the K functions as proton abstractor);
- The pair 2 acts as “steric gate” and allows only dNTPs at polymerization site and
- The pair 3 involves as “charge shielder” of dNTPs through a Mg ion (Fig. 3) and orients the α -phosphates of dNTPs for polymerization. One Mg²⁺ ion binds specifically to the beta and gamma phosphates of the dNTP as a bidentate.

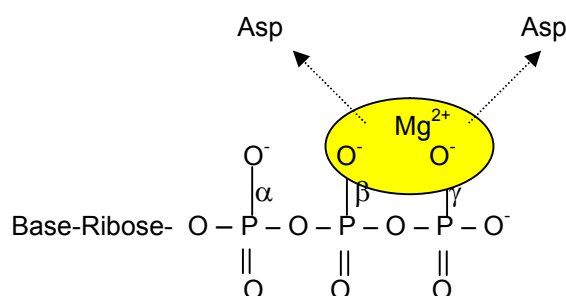


Fig. 3 dNTP-Mg ion complex where β - and γ -phosphates are shielded by the Mg ion

3.8 Functional domains of *E. coli* DNA polymerase I

As DNA polymerase I of *E. coli* is the most well studied enzyme, it is used as a model enzyme to elucidate the mechanism of action in this communication. DNA polymerase I of *E. coli* is made up of 928 amino acid residues and is a multifunctional enzyme with three different activities, viz., the polymerizing activity, the 3'→5' exonuclease activity (proof-reading function) and the 5'→3' exonuclease activity (DNA repair function). Thus, the enzyme molecule is made up of three distinct domains to perform the three different activities. These three different domains have been dissected and studied further [5].

Hans Klenow subjected the enzyme to a limited proteolysis with subtilisin, which yielded two fragments. Upon purification of the fragments by gel filtration chromatography, he found that the larger fragment, 68 kDa peptide, contained the polymerizing and proof-reading activities and the smaller fragment, 35 kDa peptide, contained the 5'→3' exonuclease activity. Further analysis of the enzyme has shown that the amino acids from 1-324 perform the 5'→3' exonuclease activity (DNA repair function); 324-517 perform the 3'→5' exonuclease activity (proof-reading function) and 517-928 perform the polymerizing function. The distance between the last two active domains is ~ 30 Å (Fig. 4).

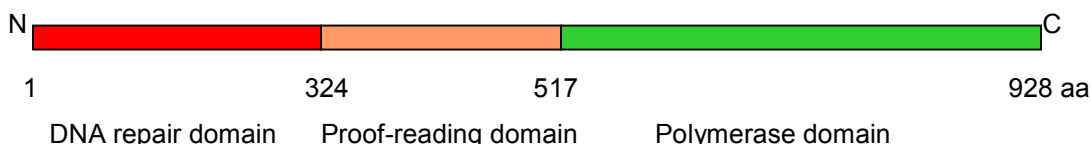


Fig. 4 Dissection of the 3 functional domains of *E. coli* DNA polymerase-1

The proof-reading site is shown to bind two metal ions and one mole of dNMP [6]. The polymerizing site consists of a deep cleft that accommodates the double stranded DNA and follows a spiral path along the DNA.

3.9 Analysis of DNA polymerase I active site

The polymerase active site was probed by a variety of techniques:

DNase foot-printing assay using DNase I and methidium-propyl EDTA-Fe²⁺ indicated that the enzyme binds to the primer terminus and covers 8 base pairs.

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. Thus, the foot printing and photo cross-linking experiment has suggested the Tyr⁷⁶⁶ in the active site [7].

However, Basu and Modak [8] who have probed the polymerase active site with pyridoxal phosphate, found a Lys⁷⁵⁸ at the active site. (Pyridoxal phosphate binds competitively to the

dNTP site through Schiff's base formation and covalently links the amino acid responsible 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 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 [9]. Zaldivar *et al.* [10] have shown that rat liver RNA polymerases I and II and yeast RNA polymerase I were also inactivated by pyridoxal phosphate and hence suggested a possible involvement of a Lysine residue in the catalytic site of RNA polymerases too.

From the above experiments, two different amino acids were implicated at the active site of the DNA polymerase I, viz., Tyr⁷⁶⁶ and Lys⁷⁵⁸. Interestingly, both the Lys and Tyr are completely conserved in some of the prokaryotic DNA polymerases analyzed by Palanivelu [11]. 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 [12].

The following observations also support Lys⁷⁵⁸ as a probable amino acid involved in the polymerization reaction rather than Tyr⁷⁶⁶:

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 [13]. Like DNA polymerases the DNA ligases are also inhibited by pyridoxal 5'-phosphate indicating the presence of a Lys at the catalytic domain of the enzyme [14]. 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 Tyr or a hydrophobic amino acid [13]. Interestingly, not only in DNA ligases, but also in RNA ligases the catalytic Lys is conserved [14].

In contrast, Tyr is found in the active site of topoisomerases, which actually involve in breaking of phosphodiester bonds. Interestingly, Tyr is conserved in all topoisomerases. In fact, the active site Tyr sometimes makes a covalent bond with the 5' end of the DNA. It is interesting to note that these topoisomerases also have highly conserved Arg/Lys, which participates in the subsequent phosphodiester bond formation [15]. Kausik *et al.* [16] and Singh and Modak [17] have shown by SDM experiments that replacing Y⁷⁶⁶ and F⁷⁷¹ by Ala, significantly affected formation of Enzyme-DNA binary complex; but most importantly the catalytic activity could not be restored in a K758A mutant.

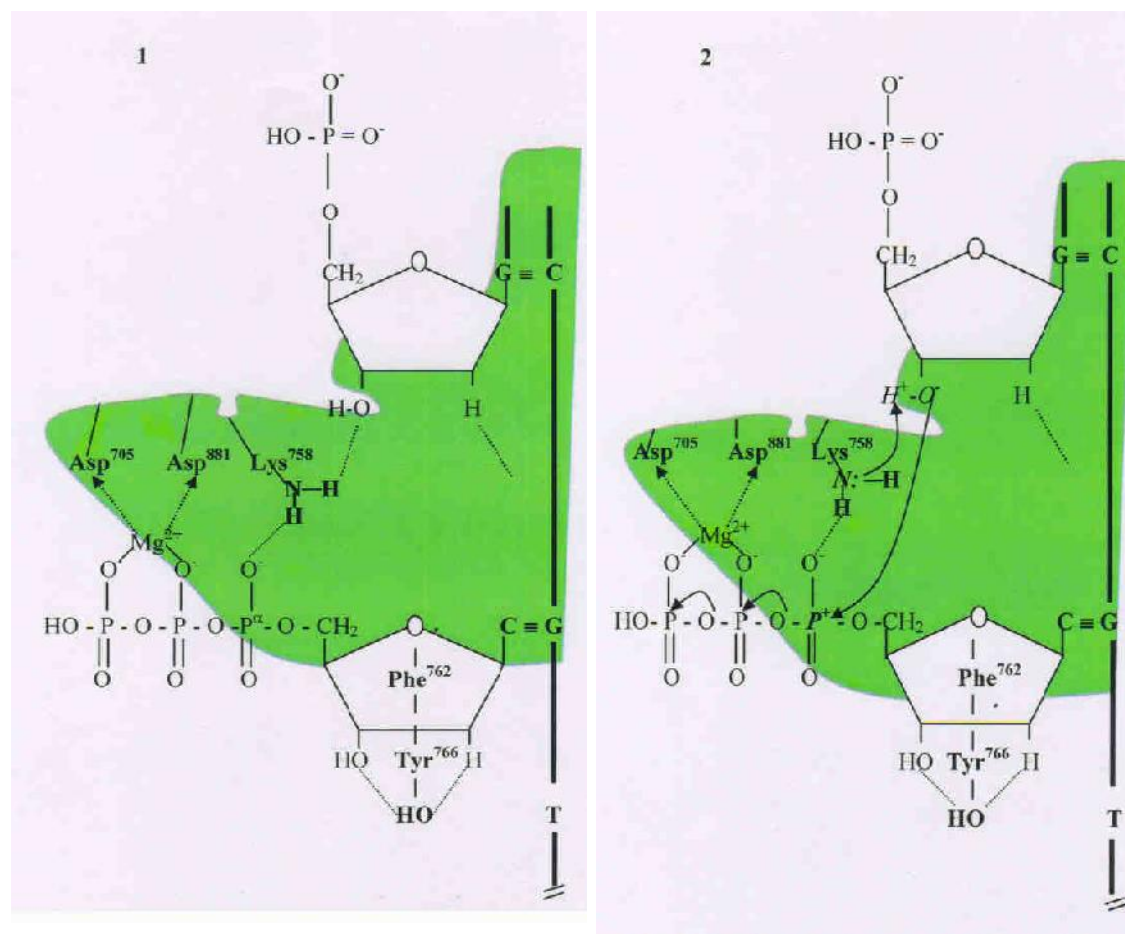
Tyr⁷⁶⁶ is suggested to be in close proximity to the 3'-OH of the primer and highly conserved in all polymerases; it could still play an important role in the polymerization function. Further analysis by SDM by Doublet and Ellenberger [18] and Astatke *et al.* [19] have shown that the critical Tyr may possibly be involved in template recognition and dNTP selection in DNA polymerases. It is also known that a highly conserved Tyr residue in reverse transcriptase controls substrate selection. In Klenow fragment, the homologous residue, Tyr⁷⁶⁶, occupies a structurally equivalent position at the C-terminus of a long helix in the fingers subdomain. It is interesting to note that the highly conserved Tyr⁹⁵⁵ residue is critical for nucleotide recognition among Family A DNA polymerases, i.e., γ polymerases from eukaryotes. Furthermore, Tyr⁹⁵⁵ is a highly conserved residue among a wide variety of DNA polymerases (Table 2). As a Family-A DNA polymerase, the γ polymerases are related to *E. coli* DNA polymerase I and bacteriophage T7 DNA polymerase, and amino acid sequence alignments reveal that Tyr⁹⁵⁵ in γ polymerases is equivalent to Tyr⁷⁶⁶ in *E. coli* pol I and Tyr⁵³⁰ in T7 DNA polymerase. Further proof of Tyr⁷⁶⁶ involvement in nucleotide selection was obtained from site directed mutagenesis; a 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 [20]. Interestingly the YG doublet is

1906 highly conserved and found to be a common pair in different types of DNA polymerases (Table
1907 2).

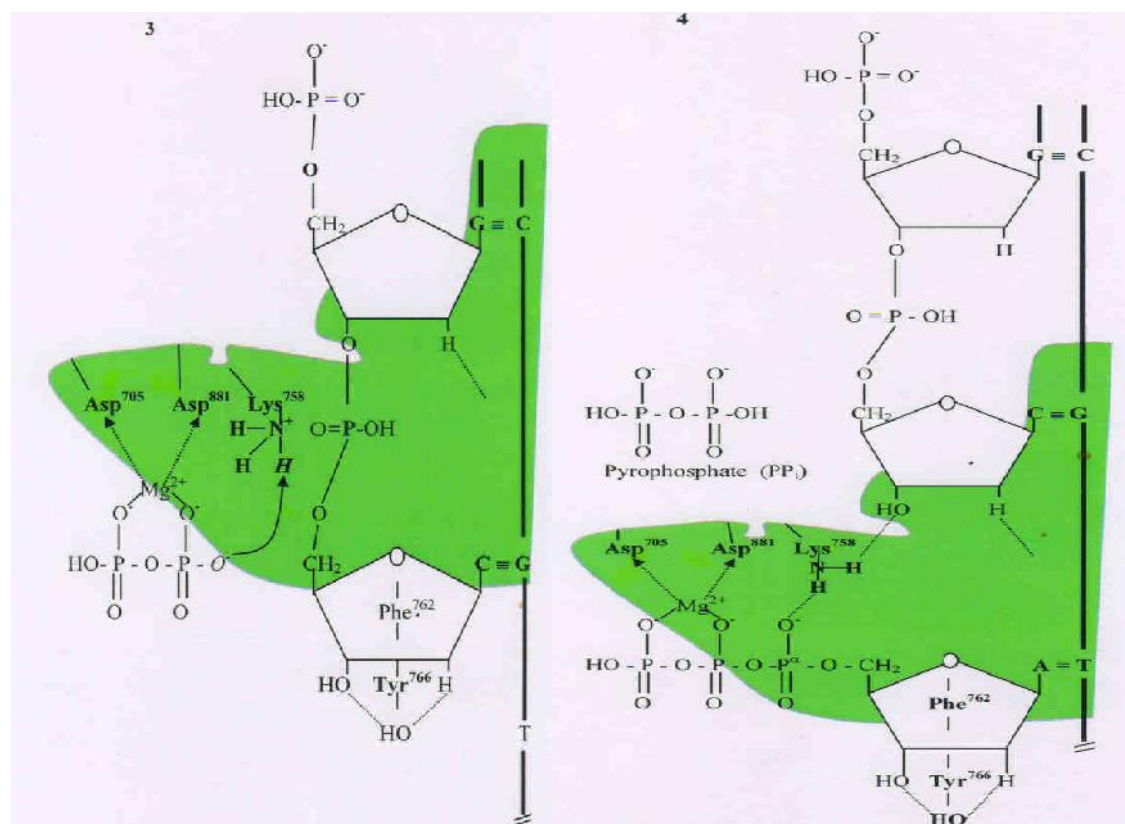
1908 Further proof is provided by crystallographic analysis of T7 DNA polymerase. The T7
1909 DNA replication complex at 2.2 Å resolution have shown that the highly conserved Lys⁵²² (=
1910 Lys⁷⁵⁸ in *E. coli* DNA pol I) actually makes contact with the α-phosphate of dNTP [21].

1911 Earlier structural, mutagenesis, and labeling studies have suggested that the incoming
1912 dNTP molecule contacts a region on one side of the polymerase cleft, primarily involving residues
1913 within the so-called "fingers" subdomain [22].

1914 Since the mechanism of action for polymerization reactions, proposed in this article is
1915 based on a proton abstraction at the catalytic site, Lys is placed as the catalytic amino acid in this
1916 communication. Thus, active site Tyr holds the complementary base possibly inserted by the
1917 finger domain onto the catalytic site and the catalytic Lys adds the dNTP to the 3'-OH of the
1918 primer terminus. The reaction essentially occurs through proton abstraction followed by a
1919 nucleophilic attack at the growing primer terminus (Figs. 5.1-5.4).
1920
1921



1922



Figs. 5.1 - 5.4 Proposed mechanism for polymerization reaction for *E. coli* DNA polymerase I

5.1. Watson-Crick base pairing of the incoming nucleotide with the template and nucleotide discrimination by steric gate amino acid tyrosine. The tyrosine-OH possibly discriminates between the 2'-H of the incoming dNTPs and 2'-OH of rNTPs. Gly/Ala residue, adjacent to the Tyr provides the flexibility necessary for the active site to change conformation to accommodate different dNTPs (not shown in the figure).

5.2. Electronic transition at the active site for proton abstraction by K followed by an electrophilic and nucleophilic attack.

5.3. Proton abstraction by the active site amino acid Lys and simultaneous formation of 3'→5' phosphodiester bond.

5.4. Transfer of the proton from Lys and formation of inorganic pyrophosphate and the next complementary nucleotide in position to get polymerized.

3.10 Analysis of the 3'-5' exonuclease active site (proof-reading function) of the DNA polymerase I

The exonuclease's proof reading site was analyzed by Derbyshire *et al.* [23, 24] by genetic, crystallographic and SDM methods. The exonuclease active site (EAS) essentially consists of a dNMP site and a metal-binding site. Therefore, dNMPs can inhibit the exonuclease reaction by product inhibition. The metal binding site consists of two subsites, viz., Subsite-A and Subsite-B and thus, EAS can bind two divalent metal ions. The subsite-A is coordinated by three

amino acids, viz., Asp³⁵⁵ - Glu³⁵⁷ - Asp⁵⁰¹ and dNMP-phosphate provides the fourth ligand. Usually a Zn atom is associated to the subsite-A.

The second metal binding site, sub-site B is mainly coordinated by Asp⁴²⁴ and possibly to a Mg atom. The subsite-B is located between dNMP-phosphate and the carboxylate of Asp⁴²⁴. The site A is very close to the 3' O- of the susceptible bond to be cleaved and the site B is very close to A. X-ray crystallographic data show that the distance between the two metal atoms is ~ 4.0 to 4.5 Å in *E. coli* (pol I) and T7 polymerases (Fig. 6).

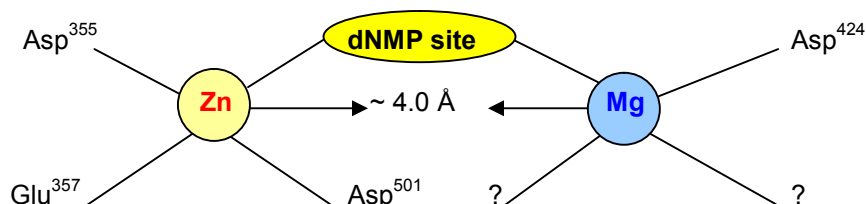


Fig. 6 Schematic diagram showing the subsites A and B of proof-reading activity of *E. coli* DNA polymerase I

The active site amino acids, which constitute the EAS, were further analyzed by site directed mutagenesis [5].

- In a double mutant with Asp³⁵⁵→Ala and Glu³⁵⁷→Ala, both the dNMP binding site and the metal binding site A were completely abolished. This mutant protein had lost the exonuclease activity, *but exhibited the polymerase activity*. This suggested that the dNMP site is coordinated to both the metal binding sites.
- In the second SDM experiment, the Asp⁴²⁴ was replaced by Ala (Asp⁴²⁴→Ala). In this mutant enzyme, the metal binding site B was abolished and exhibited no exonuclease activity. However, in this mutant protein also the polymerase activity was found to be preserved.

These two experiments suggest that the metal ion plays a direct role in proof-reading function and does not participate in polymerizing function. The SDM studies have further shown that both the metal binding sites are *functionally connected* and in the absence of one, the other cannot function. The Zn-site possibly involves in catalysis and the Mg-site linked to dNMP-phosphate could bind the dNMP site and also may provide the necessary "strain and distortion" for cleaving the susceptible bond of the wrongly added nucleotide.

3.11 Mechanism of action of proof-reading function

Polymerases occasionally make any mistake, as incorporation of even one wrong nucleotide may be detrimental to the growth and survival of organisms. In the event of a wrong nucleotide is incorporated, they cannot probably proceed further until the wrong nucleotide is removed with the help of the proof-reading function associated with them. It is just like if one tooth is bent in the zip fastener, the zip fastener cannot move further, until it is straightened. It is likely a similar situation operates in polymerases too. In such situations, the polymerase excise the wrong nucleotide by the proof-reading action, incorporates the correct nucleotide and then proceeds further.

The proof reading function is essentially explained based on Zn based catalysis. As discussed elsewhere, the exonuclease active site of *E. coli* polymerase I consists of two metal binding sites, sub-site A and B. The sub-site A is coordinated by Asp³⁵⁵, Glu³⁵⁷, Asp⁵⁰¹ and likely to a Zn atom. There is a large number of Zn containing enzymes like alkaline phosphatases, carbonic anhydrases, carboxy peptidases, thermolysin, etc. In all these enzymes, Zn plays a critical role in hydrolysis. For all catalytic Zn sites except coenzyme dependent alcohol dehydrogenase, the first two ligands are separated by a short spacer of 1 to 3 amino acids which are separated by a third ligand by a long spacer of about 20 to 120 amino acids [25, 26].

Therefore, analysis of the mechanism of action of the Zn containing enzymes will throw more light on the proof-reading function of DNA polymerases too.

Generally, in all Zn containing enzymes, a water molecule is invariably coordinated to the Zn atom as the fourth ligand, the other three ligands could be contributed by carboxylic acids (Glu, Asp), His or Cys.

During catalysis, the Zn acts as a Lewis acid and displaces the proton from the water molecule and it is held transiently by the Glu³⁵⁷, an active site amino acid at the metal binding site (the active site amino acid in other Zn containing enzymes are Glu²⁷⁰ in carboxypeptidase A; Glu¹⁴³ in thermolysin; and Glu¹¹⁷ in carbonic anhydrase) and thus the Zn becomes Zn-hydroxide (Fig. 7).

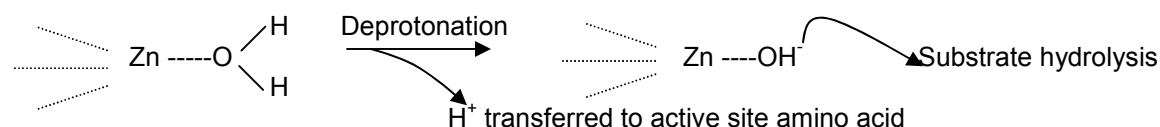
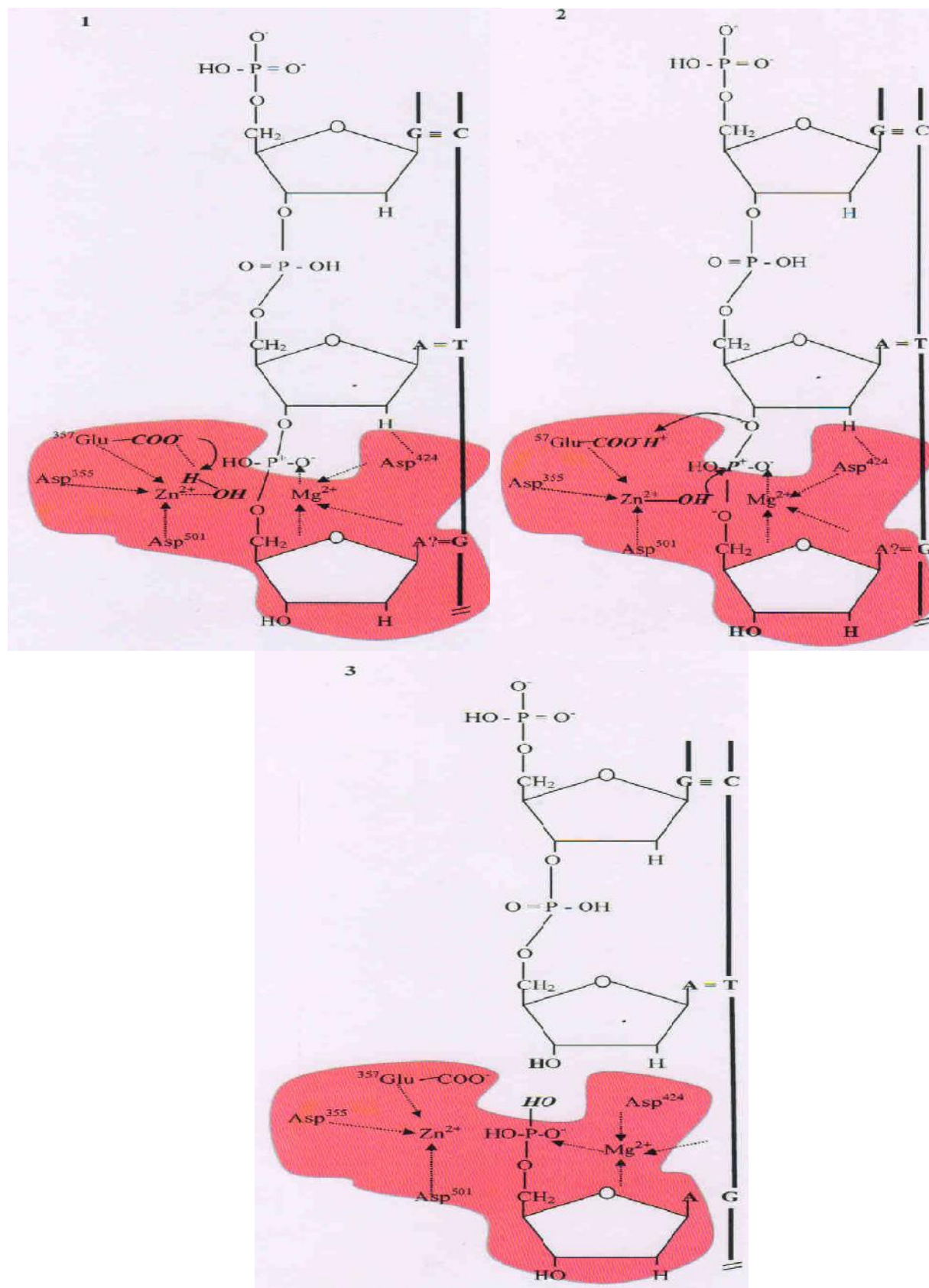


Fig. 7 Formation Zn-hydroxide during Zn mediated enzyme catalysis.

It is well known that metal hydroxides are very reactive species in chemical reactions and the Zn-hydroxide readily attacks the susceptible electrophilic center on the susceptible bond on the substrate molecule resulting in cleavage of the bond and simultaneous transfer of the proton from the active site carboxylic acid. The mechanism of 3'→ 5' exonuclease action, i.e., the proof-reading function, is also explained based on Zn-hydroxide formation and proton transfer. Derbyshire *et al.* [23] have also indicated that the nucleophilic attack of the terminal phosphodiester bond is initiated by a hydroxide ion coordinated to one of the enzyme-bound metal ions. The proposed mechanism of proof reading action is essentially based on the observations made by Beese and Steitz [6], and is illustrated in Figs. 8.1-8.3. The mismatched 3' end is a very poor substrate for further polymerization. Therefore, the 3'-5' exonuclease→ polymerase activities switch between excision and incorporation modes without dissociation of the enzyme-substrate complex [27].



Figs. 8.1- 8.3 Proposed mechanism for poof-reading function of DNA polymerases

8.1 *A wrong nucleotide is placed during polymerization (Error rate is usually one in a million).*

8.2 *Transient proton transfer to the active site amino acid Glu³⁵⁷ and Zn-hydroxide free radical formation at the proof reading active site.*

8.3 *Nucleophilic attack by Zn-hydroxide free radical at the susceptible phosphodiester bond resulting in removal of the wrongly added nucleotide with concomitant transfer of the proton from Glu³⁵⁷ to the 3' growing end. The second metal ion possibly involves in stabilizing the transient pentavalent phosphorous group during the reaction.*

Conclusion

Multiple sequence analyses have shown that a basic amino acid K/R and an YG pair are highly conserved in almost all DNA polymerases except in error prone polymerases where the YG pair is not found at the expected distance from the catalytic K/R. Site directed mutagenesis, biochemical and X-ray crystallographic analyses of DNA polymerase I from *E coli* have also suggested their involvement in catalysis and substrate binding. Based on these results, a mechanism of action is proposed for the polymerization reactions as well as for the proof reading function of DNA polymerase I from *E coli* as a model enzyme. Similar mechanism may be followed by other polymerase as the highly conserved K/R is present in all of them.

Acknowledgements

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

REFERENCES

1. Kornberg A. [1988] DNA Replication. J Biol Chem. 263: 1-4.
2. Steitz TA. [1999] DNA Polymerases: Structural Diversity and Common Mechanisms. J. biol chem. 274: 17395-17398.
3. Ohmori H, Friedberg EC, Fuchs RP, Goodman MF, Hanaoka F, *et al.* [2001] The Y-family of DNA polymerases. Mol Cell, 8: 7–8.
4. Palanivelu P. [2007] Multiple sequence analysis of polygalacturonases and invertases and phase shift in conserved motifs. Indian J Biotechnol. 6: 24-30.
5. Joyce CM, Steitz TA. [1987] DNA polymerase I: from crystal structure to function via genetics. Trends Biochem Sci. 12: 288-292.
6. Beese LS, Steitz TA. [1991] Structural basis for the 3'– 5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. EMBO J. 10: 25-33.
7. Rush J, Konisberg WH. [1990] Photoaffinity labeling of the Klenow fragment with 8-azido-dATP. J Biol Chem. 265: 4821-4827.
8. Basu A, Modak MJ. [1987] Affinity labeling of *E. coli* DNA polymerase I by pyridoxal 5'-phosphate. Biochemistry. 26: 1704-1709.
9. Monaghan N, Hay RT. [1996] Pyridoxal 5'-phosphate inhibition of adenovirus DNA polymerase. J Biol Chem. 271: 24242-24248.
10. Zaldivar MJ, Bull P, Venegas A, Valenzuela P. [1975] 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. 14: 4907-4911.
11. Palanivelu P. Enzymes, Ribozymes and DNAzymes. Twentyfirst Century Publications, Palkalai Nagar, Madurai, 2007.
12. Blasco MA, Lázaro JM, Blanco L , Salas M. [1997] 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). 94: 1619–1622.
13. Sriskanda V, Schwer B, Ho CK, Shuman S. [1999] Mutational analysis of *Escherichia coli* DNA ligase identifies 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. 27: 3953-3963.
14. Tomkinson AE, Totty NF, Ginsburg M, Lindahl T. [1991] Location of the active site for enzyme-adenylate formation in DNA ligases. Pro Natl Acad Sci (USA). 88: 400-404.
15. Redinbo MR, Stewart L, Kuhn P, Champoux JJ, Hol WGJ. [1998] Crystal Structures of Human Topoisomerase I in Covalent and Noncovalent Complexes with DNA. Science. 279: 1504-1513.
16. Kaushik N, Pandey VN, Modak MJ. [1996] Significance of the O-helix residues of *Escherichia coli* DNA polymerase I in DNA synthesis; dynamics of the dNTP binding pocket. Biochemistry. 35: 7256-66.

17. Singh K, Modak MJ. [1998] A unified DNA- and dNTP binding mode for DNA polymerases. Trends Biochem Sci. 23: 277-280.
18. Doublié S, Ellenberger T. [1998] The mechanism of action of T7 DNA polymerase. Curr Opin Struct Biol. 8: 704-712.
19. Astatke M, Grindley ND, Joyce CM. [1998] How *E. coli* DNA polymerase I (Klenow fragment) distinguishes between deoxy- and dideoxynucleotides. J Mol Biol. 278: 147-165.
20. Ponamarev MV, Longley MJ, Nguyen D, Kunkel TA, Copeland WC. [2002] Active site mutation in DNA polymerase γ associated with progressive external ophthalmoplegia causes error-prone DNA synthesis. J Biol Chem. 277: 15225-15228.
21. Doublié S, Tabor S, Long AM, Richardson, CC, Ellengerger T. [1998] Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. Nature. 391: 251-258.
22. Joyce CM. [1997] Choosing the right sugar: How polymerases select a nucleotide substrate. Proc Natl Acad Sci (USA). 94: 1619-1622.
23. Derbyshire V, Freemont PS, Sanderson MR, Beese L, Friedman JM, Joyce CM, Steitz TA. [1988] Genetic and crystallographic studies of the 3,5'-exonucleolytic site of DNA polymerase I. Science. 240: 199-201.
24. Derbyshire V, Grindley ND, Joyce CM. [1991] The 3'-5' exonuclease of DNA polymerase I of *Escherichia coli*: contribution of each amino acid at the active site to the reaction. EMBO J. 10: 17-24.
25. Vallee BL, Auld DS. [1990] Active-site zinc ligands and activated H₂O of zinc enzymes. Proc Natl Acad Sci (USA). 87: 220-224.
26. Vallee BL, Auld DS. [1992] Active zinc binding sites of zinc metalloenzymes 1: Matrix Suppl. 5-19.
27. Mizrahi V, Benkovic P, Benkovic SJ. [1986] Mechanism of DNA polymerase I: exonuclease/ polymerase activity switch and DNA sequence dependence of pyrophosphorolysis and misincorporation reactions. Proc Natl Acad Sci, (USA). 83: 5769-5773.

2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164

2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175