



DNA Polymerases: An Insight into Their Active Sites and Catalytic Mechanism

P. Palanivelu^{1*}

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

Author's contribution

*The only author performed the whole research work. Author PP wrote the first draft of the paper.
Author PP read and approved the final manuscript.*

Research Article

**Received 16th March 2013
Accepted 6th May 2013
Published 31st May 2013**

ABSTRACT

Aim: To analyze the active sites of various prokaryotic and eukaryotic DNA polymerases and propose a plausible mechanism of action for the polymerases with the *Escherichia 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 DNA 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

*Corresponding author: Email: ppmkupp@gmail.com; ppalani1953@rediffmail.com;

alpha phosphate of incoming dNTPs to the 3'-OH end of the growing primer.

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

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 *Thermus aquaticus* and *Pyrococcus furiosus* 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 localized 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' \rightarrow 3'. The overall reaction catalyzed by DNA polymerases may be written as,

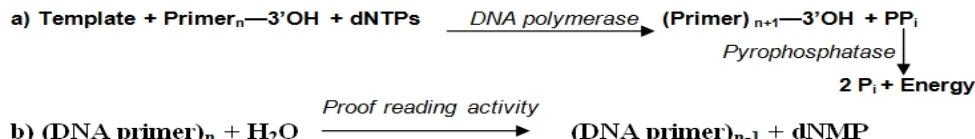


Fig. 1. Dynamics of a) DNA polymerization and b) Proof reading activity

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. MATERIAL 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

Figs. 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).

Fig. 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).

Figs. 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 Figs. 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	GKTTFRHETFQDYKGGRQQTPPELSEQFPLLRELLKAYRIPAYELDHYEADDIIGTMAAR	120
sp P00582 DPO1_ECOLI	GKGTFRDELFEHYKSHPMPDDLRAQIEPLHAMVKAMGLPLLAvgvEADDVIGTLARE	124
sp P0A550 DPO1_MYCTU	SRQTFRQLQRYPEYKANRSSTPDEFAGQIDITKEVLGALGITVLSEPGFEADDLIATLATQ	134
sp Q34996 DPO1_BACSU	GKTTFRHGTFKKEYKGGRQKTPPELSEQMFFIRELLDAYQISRYELEQYEADDIIGTIAKS	122
sp Q9F173 DPO1_SALTY	GKGTFRDELFEHYKSHPMPDDLRAQIEPLHAMVKAMGLPLLAvgvEADDVIGTLARE	124
sp P59200 DPO1_STRR6	GKTTFRTEMADYKGGRAKTPDFREQFFIRELLDHMGIRHYELAQYEADDIIGTLDKL	125
sp O32801 DPO1_LACLM	GKTTFRTEMADYKGGRSCKTPDFREQFFIRELLDHMGIRHYELAQYEADDIIGTLDKM	126
sp P43741 DPO1_HAEIN	GKGTFRDEMFEQYKSHRPPMPDDLRKQIQLHDMIRALGIPLLVVEGIEADDVIGTLALQ	124
sp Q9ZJE9 DPO1_HELPJ	QTKTAKRAEKLGEYKQNPKDAPKEMLLQIPIALEWLQRMGFTCVEVGGEADDVIASLATL	130
sp Q9CDS1 DPO1_LACLA	GKTTFRTEMADYKGGRSCKTPDFREQFLPIKEMIEKLGIRHYELAQYEADDIIGTLDKM	126
sp P59199 DPO1_STRPN	GKTTFRTEMADYKGGRAKTPDFREQFFIRELLDHMGIRHYELAQYEADDIIGTLDKL	125
sp Q9S1G2 DPO1_RHILE	SAKTFRKDLYDAYKANRSAPPEELIQPGLIREATRAFNLPCIETEGEADDIIATYARQ	147
sp Q92GB7 DPO1_RICCN	GGKNFRHQIYPDYKANRPPPPEDLITQLPLVRDVASNLFNPILEKNGYEADDIIATPATK	122
sp P0A551 DPO1_MYCBO	SQTFRLQRYPEYKANRSSTPDEFAGQIDITKEVLGALGITVLSEPGFEADDLIATLATQ	134
sp Q9HT80 DPO1_PSEAE	KGPTFRDELFAEYKANRSPMPDDLRVQVPELHASVRAVLGLPLLCVEGVEADDVIGTLARS	121
sp Q04957 DPO1_BACCA	GKTTFRHEAFQEYKGRQQTPPELSEQFPLLRELLRAYRIPAYELENYEADDIIGTIAAR	120
sp P46835 DPO1_MYCLE	SRKTFRSECYAGYKANRSSIPAEEFHGQIDITKEVLGALGITVFAEAGEADDLIATLATQ	135
sp P52027 DPO1_DEIRA	PVKTFRHEQYEGYKSGRAQTPELPGQINRIRALVDALGFPRLLEPGYEADDVIASLTRM	163
	.	*
	*	*
	*	*
	:	:
	*	*
	:	:
	*****	:..:

sp P52026 DPO1_BACST	AER--EGFAVKVISGDRDLTQLASPQVTVEITKKGITDIESYTPTVVVKYGLTPEQIVD	178
sp P00582 DPO1_ECOLI	AEK--AGRPLVLISTGDKDMAQLVTNPNTLINTMTN---TILGPPEVVNKYGVPPPELIID	178
sp P0A550 DPO1_MYCTU	AEN--EGYRVLVVTGCDRDLAQVLSDDVTVLYPRKGVSELTRFTPEAVVEKYGLTPRQYPD	192
sp Q34996 DPO1_BACSU	AEK--DGFEVKVSGDKDLTQLATDKTTVAITRKGITDVEFYTPHEVKEKYGLTPEQIID	180
sp Q9F173 DPO1_SALTY	AEK--VGRPLVLISTGDKDMAQLVTNPNTLINTMTN---TILGPDEVVNKYGVPPPELIID	178
sp P59200 DPO1_STRR6	AEQ--DGFDTITIVSGDKDLIQLTDEHTVVEISKKGVAEFEAFTPDPYIMEEMGLTPAQFID	183
sp O32801 DPO1_LACLM	AEAPDVFNDVTIVTCGDKDMIQLVLDGNTRVEISKKGVAEFEETPDYIMEEMGLTPSQFID	186
sp P43741 DPO1_HAEIN	ASS--LGKKVLISTGDKDMAQLVDDNIMLINTMNN---SLLDRKGVIEKYGIPPELIID	178
sp Q9ZJE9 DPO1_HELPJ	SP----YKTRIYSKDKDFNQLLSKIALFDGK----TEFLAKDCVEKYGILPSQFTD	179
sp Q9CDS1 DPO1_LACLA	AEA--PNVNFDTIVTCGDKDMIQLVLDGNTRVEISKKGVAEFEETPDYIMEEMGLTPAQFID	186
sp P59199 DPO1_STRPN	AEQ--DGFDTITIVSGDKDLIQLTDEHTVVEISKKGVAEFEAFTPDPYIMEEMGLTPAQFID	183
sp Q9S1G2 DPO1_RHILE	AEA--TGADVTIVSSDKDLMQLVSPNVHMYDSMKD---KQIGIPDVIEKWGVPPPEKMD	201
sp Q92GB7 DPO1_RICCN	TAA--LGAHVVISSDKDLMQLMTENIKIYDPLKG---KYITEDVVKKFGTTSDLR	176
sp P0A551 DPO1_MYCBO	AEN--EGYRVLVVTGCDRDLAQVLSDDVTVLYPRKGVSELTRFTPEAVVEKYGLTPRQYPD	192
sp Q9HT80 DPO1_PSEAE	SAA--ADRPVVIStGDKDMAQLVGDGHITLVNTMTG---SRLDVDGVKEKFGVGPELIID	175
sp Q04957 DPO1_BACCA	AEQ--EGFEVKVISGDRDLTQLASPHTVVDITKKGITDIEPYTPEAVREKYGLTPEQIVD	178
sp P46835 DPO1_MYCLE	AEN--EGYRVLVVTGCDRDLAQVLVSNDTVLYPRKGVSELTRFTPEAVIEKYGVTPAQYPD	193
sp P52027 DPO1_DEIRA	AEG--KGYEVIRVTSDRDAYQLLDEHVKVIAN---DFSLIGPAQVEEKYGVTVRQWVD	216
	:	:
	: * : * :	** . : :
		: * : :

sp P52026 DPO1_BACST	NMEFTGVKVDTKRLEQMGAELTEQLQAVERRIYELAGQEFNIN	SPKQQLGTVLFDKLQLPV	546								
sp P00582 DPO1_ECOLI	RIERNGVKIDPKVLHNHSEELTLRLAEEKHAHEIAGEEFNL	STPKQQLQTILFEKGQGIKP	598								
sp P0A550 DPO1_MYCTU	KMESAGIAVDLPLMTELQSQFDQIRDAAEAYVGIVKQINL	GSPKQQLGVVLFDELGMPK	573								
sp O34996 DPO1_BACSU	EMESTGVKVVDLRLKRMGEELGAKLKEYEKIHEIAGEEPFN	INSPKQQLGVILFEKGIGLPV	550								
sp Q9F173 DPO1_SALTY	RVERNGVKIDPAVLHKHSSEEITLRLAEELEKKAHDIAGEAFN	LNSSTPKQQLQTILFEKGQGIKP	598								
sp P59200 DPO1_STRR6	KMEIAGIVVVKETLLEMQAENELVIEKL	TQEIYELAGEEFNVNSPKQQLGVLLFEKLGLPL	545								
sp O32801 DPO1_LACLM	KMEITGISVSQNTLEEIGAENEELKASLTREIYDLAGEEF	FNNSPKQQLGVILFEKLQLPV	548								
sp P43741 DPO1_HAEIN	RMERVGVLIDSDALFMQSNEIASRLTALEKQAYALAGQF	NLASTPKQQLQEILDKLELPV	601								
sp Q9ZJE9 DPO1_HELPJ	GMEFGFKIDAPYFKRLEQEFKNEVLNRQILD	LIGVDFNLNSPKQQLGEVLYDKLGLPK	575								
sp Q9CDS1 DPO1_LACLA	KMEITGIAVSQNTLEEIGAENEELKASLTREIYDLAGEEF	FNVNSPKQQLGVILFEKLQLPV	548								
sp P59199 DPO1_STRPN	KMEIAGIMVKKETLLEMQAENELVIEKL	TQEIYELAGEEFNVNSPKQQLGVLLFEKLGLPL	545								
sp Q9S1G2 DPO1_RHILE	RMEARGITVDQRLSRLSGELAQGAARLEDEIYVLAGER	FVNIGSPKQQLGDILFGKMGLSG	679								
sp Q92GB7 DPO1_RICCN	KMEKVGITVDANYLQKLSAEFGTEILKLEE	EIFALSCTFKNIGSPKQQLGEILFEKMQLPF	545								
sp P0A551 DPO1_MYCBO	KMESAGIAVDLPMTELQSQFDQIRDAAEAYVGIVKQINL	GSPKQQLGVVLFDELGMPK	573								
sp Q9HT80 DPO1_PSEAE	RIERNGVALVDANLLG1QISRELGEKMLAVERQAYDLAG	OEFNLGSPKQQLQIALYDKLGLPL	582								
sp Q04957 DPO1_BACCA	EMEFAGVKVDTKRLEQMGEELAEQLRTVEQRIYELAGQEF	NINSPKQQLGVILFEKLQLPV	546								
sp P46835 DPO1_MYLE	DMEKAGIAADLRLLT	ELQSQFDQIRDAAEAYAVIGKQINLSSPKQQLGVVLFEE	LGMPK	577							
sp P52027 DPO1_DEIR	RMEVRGVQVQDSFLQTL	SQIAGVRLADLESQIHEYAGEEFHISPKQQLBTVLYD	KLELAS	627							
	:	*	****	***:	*: :	:	****	*: :	** :	** :	:

sp P52026 DPO1_BACST	TGKVHTMFNQALTQ	GRLSSVEPNLQNI	IRLEEGRKIRQAFVPSEP--DW-LI	FAADYS	655							
sp P00582 DPO1_ECOLI	TGRVHTSYHQAVT	AATGRLSSTD	PNLQNI	EVRRNEGRRIHQAFIAPE---DY-VIVSADYS	707							
sp P0A550 DPO1_MYCTU	DGRIHTTFNQ	QTIAATGRLS	STEPNLQNI	IIRTDA	RRIRDAFVVG---GYAELMTADYS	682						
sp O34996 DPO1_BACSU	SHKVHTRF	NQALTQ	GRLSSTD	PNLQNI	HIRLEEGRKIRQAFVPSEK--DW-LI	FAADYS	659					
sp Q9F173 DPO1_SALTY	TGRVHTSYHQAVT	AATGRLSSTD	PNLQNI	EVRN	EEGRRIHQAFIAPE---DY-LIVSADYS	707						
sp P59200 DPO1_STRR6	DGKIHTRYVQD	LQTQ	GRLSSVDP	PNLQNI	IARLEQG	RKIRKA	VPEWE--DS-VL	LSSDYS	655			
sp O32801 DPO1_LACLM	DGKIHTRYVQD	LTQGRLSSVDP	PNLQNI	EVRL	EEGRKIRKAFVPS-Q--DS	LLLSSDYS	655					
sp P43741 DPO1_HAEIN	TGRVHTSYHQAVT	AATGRLSSSD	PNLQNI	I	IRNEEGRHRHQAFIARE---GY-SIVA	ADYS	710					
sp Q9ZJE9 DPO1_HELPJ	DDKIHTTFI	QTGTAT	GRLSSSD	PNLQNI	I	VRSPKG	LLRKGFIASSK--EY-CLLG	VADYS	679			
sp Q9CDS1 DPO1_LACLA	DKGIHTRYVQ	DLTQ	GRLSSVDP	PNLQNI	I	EVRL	EEGRKIRKAFVPS-K--DS	LLLSSDYS	655			
sp P59199 DPO1_STRPN	DKGIHTRYVQ	DLTQ	GRLSSVDP	PNLQNI	I	ARLEQG	RKIRKA	VPEWE--DS-VL	LSSDYS	655		
sp Q9S1G2 DPO1_RHILE	TKRVHTSY	SLAS	TTTGRLSS	SE	PNLQNI	I	V	TAEGRKR	I	ARTAFISTP--GH-KLISADYS	789	
sp Q92GB7 DPO1_RICCN	THRVTHTFL	QT	GRLSSQEP	PNLQNV	I	IRSSEG	NQKIRKAFIAE---	GY-KLISADYS	656			
sp P0A551 DPO1_MYCBO	DGRIHTTFN	QTIAAT	GRLS	STE	PNLQNI	I	I	RTDA	RIRDAFVVG	---GYAELMTADYS	682	
sp Q9HT80 DPO1_PSEAE	TGRVHTSYHQAV	AATGRLSSSD	PNLQNI	I	I	TAEGR	RHQAFV	VAPQ---GY-KLISADYS	692			
sp Q04957 DPO1_BACCA	TKVHTIF	NQALTQ	GRLS	STE	PNLQNI	I	I	LEEGRK	I	RQAFVPSES--DW-LI	FAADYS	656
sp P46835 DPO1_MYLE	DGRIHTTFN	QTIAAT	GRLS	STE	PNLQNI	I	I	RTNA	GRQIRD	AFVVGSENN	GYTELMTADYS	689
sp P52027 DPO1_DEIR	TGRLH	TTFAQT	AVATG	RGLSS	I	PNLQNI	I	I	SELGR	EIRKGF	IAED---GF-TLIAADYS	737
	:***:	.	*****	.*****:	*	*	*	**	**	*	.	***

sp P52026 DPO1_BACST	QIELRVL	LAH	AEDDN	LIEAFRRGLDI	HTKTAM	DFHV	SE-EDV	TANMR	RQAKAVNFG	IFYV	714										
sp P00582 DPO1_ECOLI	QIELR	IMAHL	SRD	KGLLTA	FAEG	KD	IH	RATAAE	EVFG	LPPL-ETVT	SEQR	RSAKAINFG	LIY	766							
sp P0A550 DPO1_MYCTU	QIELR	IMAHL	SGD	EGLIEA	FTN	GEDL	H	HFVAS	RAFG	VI	-DEVT	GELRRV	KAMS	YGLAY	741						
sp O34996 DPO1_BACSU	QIELRVL	LAH	SKDEN	DLIEA	FTND	MDI	I	HTKTAM	DFHV	VAK	-DE	VTSEQR	RQAKAVNFG	IFYV	718						
sp Q9F173 DPO1_SALTY	QIELR	IMAHL	SRD	KGLLTA	FAEG	KDIH	RATAAE	EVFG	LPPL-	DSVT	GELRRV	QKRS	AKAINFG	LIY	766						
sp P59200 DPO1_STRR6	QIELRVL	LAH	SKDEH	L	LI	KA	QEG	AD	IHTS	TA	MRVFG	I	ERPDNV	TAN	RNA	KAVNFGV	YY	715			
sp O32801 DPO1_LACLM	QIELRVL	LAH	SADEH	L	LI	DAF	KHG	A	DI	H	T	STAM	RVFG	I	KA	VNFGV	YY	715			
sp P43741 DPO1_HAEIN	QIELR	IMAHL	SGDQGL	QLINAFS	TSQ	GKDIH	STA	AE	EVFG	LPPL-	DSV	TGEL	RRV	SEQR	RNA	KAINFG	YY	769			
sp Q9ZJE9 DPO1_HELPJ	QIELR	LLAHL	FSQDK	KDLM	EAFL	KGR	DIH	L	ETSK	KALFG	---	EDLAKE	KRS	KS	INFGLV	Y	735				
sp Q9CDS1 DPO1_LACLA	QIELRVL	LAH	SGD	DEHL	I	DAF	KHG	A	DI	H	T	STAM	RVFG	I	KA	VNFGV	YY	715			
sp P59199 DPO1_STRPN	QIELRVL	LAH	SGD	DEHL	I	DAF	KHG	A	DI	H	T	STAM	RVFG	I	KA	VNFGV	YY	715			
sp Q9S1G2 DPO1_RHILE	QIELRVL	LAH	SGD	DEHL	I	DAF	KHG	A	DI	H	T	STAM	RVFG	I	KA	VNFGV	YY	715			
sp Q92GB7 DPO1_RICCN	QIELR	ILAHL	AN	I	DALK	QAF	KAF	Q	FIN	KDDI	HT	TC	ACQ	I	FN	LNQ	KAINFG	IY	715		
sp P0A551 DPO1_MYCBO	QIELR	IMAHL	SGD	DEGL	I	EAF	N	G	FTN	GEDL	H	HFVAS	RAFG	VI	-DE	VTGELRRV	KAMS	YGLAY	741		
sp Q9HT80 DPO1_PSEAE	QIELR	IMAHL	AK	ADDL	M	QD	LL	I	FR	HD	D	V	H	RATAA	EVFG	VPL	-EDVSGD	QKRS	AKAINFG	LIY	751
sp Q04957 DPO1_BACCA	QIELR	VL	AH	ADDNL	MEA	FRR	LDI	H	T	AM	D	FQV	SE-	DEV	T	PNM	R	QKRS	AKAINFG	IFYV	715
sp P46835 DPO1_MYLE	QIELR	LLAHL	ADDPL	MQQAF	VEG	GAD	I	H	R	T	A	QVL	GLDE-	ATV	DAN	QRR	A	KTVNFGV	LY	796	
sp P52027 DPO1_DEIR	*****:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	*****	:	*****:	*	*	

sp P52026 DPO1_BACST sp P00582 DPO1_ECOLI sp P0A550 DPO1_MYCTU sp O34996 DPO1_BACSU sp Q9F173 DPO1_SALTY sp P59200 DPO1_STRR6 sp O32801 DPO1_LACLM sp P43741 DPO1_HAEIN sp Q9ZJE9 DPO1_HELPJ sp Q9CDS1 DPO1_LACLA sp P59199 DPO1_STRPN sp Q9S1G2 DPO1_RHILE sp Q92GB7 DPO1_RICCN sp P0A551 DPO1_MYCBO sp Q9HT80 DPO1_PSEAE sp Q04957 DPO1_BACCA sp P46835 DPO1_MYCLE sp P52027 DPO1_DEIRA	GISDYGLAQNLNITRKEAAEFIGYFASFPGVKQYMDNIVQEAKQKGYVTTLLHRRRYLP 774 GMSAFGLARQLNIPRKEAQKYMDFYFERYPGVLEYMERTRAQAKEQGYVETLDGRRLYLP 826 GLSAYGLSQQLKISTEEANEQMDAYFARFGGVRDYLRAVERARKDGYTSTVLGRRRYLP 801 GISDYGLSQNLGITRKEAGAFIDRYLESFQGVKAYMEDSVQEAKQKGYVTTLMHRRRYIP 778 GMSAFGLSRQLNIPRKEAQKYMDFYFERYPGVLEYMERTRAQAKEQGYVETLEGRRLYLP 826 GISDFGLSNNLGISRKEAKAYIDTYFERFPGIKNYMDDEVREARDKGYVETLFKRRREL 775 GISDFGLARNLNLGIRAKYMDLYFERYPGIKTYMENIVREARDKGFVETMSHRRRKIP 775 GMSAFGLSRQLGISRDAQKYMDFYFQYPSVQQFMTDIREKAKAQGYVETLFGRRLYLP 829 GMGSKKLSETLSIPLSEAKSYIEAYFKRFPISKDYLNGMREEILKTSKAFTLLGRYRV-F 794 GISDFGLARNLNLGIRTRKDAKNYIETYFERYPGIKTYMENIVREARDKGFVETMSHRRRKIP 775 GISDFGLSNNLGISRKEAKAYIDTYFERFPGIKNYMDDEVREARDKGYVETLFKRRREL 775 GISAFGLANQLSIERSEAGDYIKKYFERFPGIKDYMESRKAMARDKGYVETIFGRINYP 908 GISAFGLAKQLNVNTGEASEYIKKYFAEYKGVQEYMEQTAKAFASSNGYVTCFGRKCFVP 775 GMSAFGLSQQLKISTEEANEQMDAYFARFGGVRDYLRAVERARKDGYTSTVLGRRRYLP 801 GMSAFGLAKQIGVERKEAQAYIDRYFARYPGVLAYMERTRAQAAEQGFVETLFGRRLYLP 811 GISDYGLAQNLNISRKEEAAEFIGYFESFPGVKRYMENIVQEAKQKGYVTTLLHRRRYLP 775 GLSAYGLATQLKISTEEAKLQMEQYFARFGGVRDYLMDVVEQARKDGYTSTVLGRRRYLP 808 GMSAHRLSNDLGIPYAEAATFIEIYFATYPGIRRINHTLDFGRTHGYVETLYGRRRYVP 856 * . * : : * : * : : . : * : : . : : *
sp P52026 DPO1_BACST sp P00582 DPO1_ECOLI sp P0A550 DPO1_MYCTU sp O34996 DPO1_BACSU sp Q9F173 DPO1_SALTY sp P59200 DPO1_STRR6 sp O32801 DPO1_LACLM sp P43741 DPO1_HAEIN sp Q9ZJE9 DPO1_HELPJ sp Q9CDS1 DPO1_LACLA sp P59199 DPO1_STRPN sp Q9S1G2 DPO1_RHILE sp Q92GB7 DPO1_RICCN sp P0A551 DPO1_MYCBO sp Q9HT80 DPO1_PSEAE sp Q04957 DPO1_BACCA sp P46835 DPO1_MYCLE sp P52027 DPO1_DEIRA	DTTSRNFNRVRSFAERTAMNTPIQGSADIICKAMIDLSQLVLRERL--QARILLQVHDEL 832 DIKSSNGARAAAERAANAPMQGTAADIICKRAMIAVDALQAEQP--RVRMIMQVHDEL 884 ELDSSNRQVREAAERAALNAPIQGSAAIDIKVAMIQVDKALNEAQL--ASRMILLQVHDEL 859 ELTSRNFNRVRSFAERTAMNTPIQGSAAIDIICKAMIDMAAKLKEKQL--KARILLQVHDEL 836 DIKSSNAARRAGERAARAINAPMQGTAADIICKRAMIAVDALQAEQP--RVRMIMQVHDEL 884 DINSRNFNIRGFAERTAINSPPIQGSAAIDLKIAMIQLDKALVAGGG--QTKMLLQVHDEI 833 DINARNFNVRGFAERTAINSPPIQGSAAIDLKIAMIIDLKALTERQS--KSLLLQVHDEI 833 DINSSNMRKGAAERVAINAPMQGTAADIICKRAMIKLDEVRH-DP--DIEMIMQVHDEL 886 DFTGVNDYVKGNYLREGVNAIPQGSASDLKLGLMLKVSERFKN-NP--SVRLLLQVHDEL 851 DINARNFNVRGFAERTAINSPPIQGSAAIDLKIAMIIDLKALSARDF--KSLLLQVHDEI 833 DINSRNFNIRGFAERTAINSPPIQGSAAIDLKIAMIQLDKALVAGGG--QTKMLLQVHDEI 833 EIRSSNPSVRAFNERAARAINAPPIQGSAAADVIRRAMIKIEPALVEVGLADRVRMILLQVHDEL 968 LIHDK--KLKQFAERAARAINAPPIQGTNTDIIKIAMINLDQIEKRKL--KTRLVLQIHDEL 831 ELDSSNRQVREAAERAALNAPIQGSAAIDIKVAMIQVDKALNEAQL--ASRMILLQVHDEL 859 EIHSKNGAMRKAARTAINAPMQGTAADIMKRAMAVDNWLQESGL--DARVILQVHDEL 869 DTTSRNFNRVRSFAERMAMNTPIQGSAAIDIICKAMIDLNLARKEERL--QARILLQVHDEL 833 ELDSSNRQIREAAERAALNAPIQGSAAIDIKVAMIAVDKSLKQAKL--ASRMILLQVHDEL 866 GLSSRNVRQREAAERLAYNMPPIQGTAADIMKLMQVQLDPQLDA--I--GARMILLQVHDEL 912 : : * . * : * : * : * : : . : : * : * : * : *
sp P52026 DPO1_BACST sp P00582 DPO1_ECOLI sp P0A550 DPO1_MYCTU sp O34996 DPO1_BACSU sp Q9F173 DPO1_SALTY sp P59200 DPO1_STRR6 sp O32801 DPO1_LACLM sp P43741 DPO1_HAEIN sp Q9ZJE9 DPO1_HELPJ sp Q9CDS1 DPO1_LACLA sp P59199 DPO1_STRPN sp Q9S1G2 DPO1_RHILE sp Q92GB7 DPO1_RICCN sp P0A551 DPO1_MYCBO sp Q9HT80 DPO1_PSEAE sp Q04957 DPO1_BACCA sp P46835 DPO1_MYCLE sp P52027 DPO1_DEIRA	ILEAPKEEIERLCRLVPEVME-QA---VTLRVPLKVDYHYGPTWYD-AK 876 VFEVHKDDVDAAVAKQIHLQME-NC---TRLDVPLLVEVGSGENWDQ-AH 928 LFEAPGERERVEALVRDKMG-GA---YPLDVPLLEVSGYGRSWDAAH 904 IFEAPKEEIEILEKLVPEVME-HA---LALDVPLKVDFFASGPSWYD-AK 880 VFEVHKDDLDAAVAKRHLQME-NC---TRIDVPLLVEVGSGENWDQ-AH 928 VLEVPKSELVEMKKLVQTM-EA---IQLSVPLIADENEGATWYE-AK 877 ILDVPLEELEDIKALVQTM-EA---IELAVPLKVDNTGKTWYE-AK 877 VFEVSEKVAFFREIQKQHME-AA---AELVVPLIVEVGVGQNWE-AH 930 IFEIEEKNAPELQQEIQRILNDEV---YPLRVPLSETSFAFIKRWNELKG 897 ILDVPLEELDEIKVLVQTM-EA---IELAVPLKVDENTGKTWYE-AK 877 VLEVPKSELVEMKKLVQTM-EA---IQLSVPLIADENEGATWYE-AK 877 IFEVEDQDVEKAMPVIVSVM-EATMPALEMRVPLRDARAATNWDE-AH 1016 LFEAPIDEVEITPIKKIME-NS---TNMVAIPITIERRAGNNWME-IH 875 LFEAPGERERVEALVRDKMG-GA---YPLDVPLLEVSGYGRSWDAAH 904 VLEVRREDLVEQVCEGIRPLMS-GA---ATLDVPLLVEAGVGSNWDE-AH 913 ILEAPKEEMERLCRLVPEVME-QA---VTLRVPLKVDYHYGSTWYD-AK 877 LFEVAIGEREQIEAMVREQM-G-SA---YPLDVPLLEVSGFGRSWGAAAH 911 LIEAPLDKAEQVAALKVME-NV---VQLKVPLAVEVTGPWFD-TK 956 : : * : * : * : *

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.

Fig. 2B Prokaryotic DNA polymerases (pol II)

sp Q9V2F3 DP2S_PYRAB	LISDIHVGSRFCCEKAFLKFLWLNHGVESKEEEEIVSRVKYLTIAGDVVDGIGIYPGQY	416
sp Q9HMR7 DP2S_HALSA	LISDVHVGSQEFAADAWRSFADWLHT-----PAAESVEYLLIAGDMVEGVGVYPGQD	308
sp Q9HLK5 DP2S_THEAC	SISDIHVGSKTFRKNEFEAMVRWISGS-----DPDASRVVKYLILSGDVVDGIGVYPDQE	288
sp O27456 DP2S_METTH	FISDVHIGSQTFLEDAFMKFKWINGDFGSEEQRSLAADVVKYLVVAGDIVDGIGIYPGQE	285
sp B0R7U1 DP2S_HALS3	LISDVHVGSQEFAADAWRSFADWLHT-----PAAESVEYLLIAGDMVEGVGVYPGQD	308
sp O28484 DP2S_ARCFU	FLSDTHFGSKFLEKEWEMFVRWLKGEGVGGKKSQNLAEKVKYIVIAGDIVDGIGVYPGQE	290
sp P81412 DP2S_PYRFU	LISDIHVGSKFECENAFIKFLEWLNGNVTKEEEEIVSRVKYLIIAGDVVDGVGVYPGQY	410
sp Q58113 DP2S_METJA	FLSDIHVGSKFELKEFKIRFLNGDWDNELEEKVVSRLKYICIAGLDVGDGVGVYPGQE	386
	:** * .***: * .. : : .. . :*: : :*: :*: :*: :*. *	
sp Q9V2F3 DP2S_PYRAB	SDLVIIPDIFDQYEALANLLANVPEHITMFIPGNHDAARPAIPQPPEFYKEYAKPIYKLKN	476
sp Q9HMR7 DP2S_HALSA	EELDIVDIYDQYETFAEHLKDVPGDMIEVMIPGNHDAVRЛАЕРQPQAFDEEL-RSIMRAHD	367
sp Q9HLK5 DP2S_THEAC	NDELIINPLEQYANLAEYLVDVPEDVKVFMPGNHDITVRLSEPPQPVFPCKI-RDLFEP-N	346
sp O27456 DP2S_METTH	KELLIRDIHEQYEEAARLFGDIRSDIKIVMIPGNHDSRIAEPQPAPIPEEYAKSLYSIRN	345
sp B0R7U1 DP2S_HALS3	EELDIVDIYDQYETFAEHLKDVPGDMIEVMIPGNHDAVRЛАЕРQPQAFDEEL-RSIMRAHD	367
sp O28484 DP2S_ARCFU	DDLAISDIYGQYEFASHLDEIPKEIKIIVSPGNHDAVRQAEQPQAFGEI-RSLFP-KN	348
sp P81412 DP2S_PYRFU	ADLTIPDIFDQYEALANLLSHVPKHTMFIAPGNHDAARQAIPQPPEFYKEYAKPIYKLKN	470
sp Q58113 DP2S_METJA	EDLYEVVDIIEQYREIAMYLDQIPEHISIIISPGNHDAVRPAEPQPQPKLPEKI-TKLFNRDN	445
	:* : ** * : .. .: .. :*****: * : *** : : : :	
sp Q9V2F3 DP2S_PYRAB	AIIISNPAPAVIRLHGRDFLIAHGRGIEDVVSFVPG--LTHHKPGLPMVELLMKRHLAPTFG	534
sp Q9HMR7 DP2S_HALSA	ARITSNPSTVTIDGVSVLLYHGVSLLDEVIAEHPSDDVTYDDPQNAMELLKKRHVAPPFG	427
sp Q9HLK5 DP2S_THEAC	VAFLPNPYNLKLEGKNVLVYHGMISLNDMIELIPG--ANYDSIGKAIEAILVRKRHLSPKG	404
sp O27456 DP2S_METTH	IEFLSNPSLVSLDGVRTLIIYHGRSFDDMAMSNG--LSHERSDLIMEELLEKRHLAPIYG	403
sp B0R7U1 DP2S_HALS3	ARITSNPSTVTIDGVSVLLYHGVSLLDEVIAEHPSDDVTYDDPQNAMELLKKRHVAPPFG	427
sp O28484 DP2S_ARCFU	VEHVGNPAYVDIEGVVKVLIIYHGRSIDDIISKIPR--LSYDEPQKVMEELLKRHLSPKG	406
sp P81412 DP2S_PYRFU	AVIIISNPAPAVIRLHGRDFLIAHGRGIEDVVSFVPG--LTHHKPGLPMVELLMKRHVAPMFG	528
sp Q58113 DP2S_METJA	IYFVGNPCTLNIGHGFDLTLHYHGRSFDDLVGQIRA--ASYENPVTIMKELIKRRLLCPTYG	503
	** : :.* : * : ** ..::: .. .: .. : : : * . * . *	
sp Q9V2F3 DP2S_PYRAB	GKVPiapdpedllvievvpdlnqmgvhvyydavvyrgvqlvnnsatwqaqtefqkmvnivp	594
sp Q9HMR7 DP2S_HALSA	GRTRLAPEAEDHLAIDTVPDFHTGHVHKLGVGIHHNVRIVNSGCWQHQTAQFQESVNISP	487
sp Q9HLK5 DP2S_THEAC	GNTPMIPSADVYHvievpdiftGHIIHSHYIGNYKGVRVYVNSSTWQSQTEYQKMMNFNP	464
sp O27456 DP2S_METTH	ERTPLASEIEDHLVIDEVPVLHTGHVHINAYKKYKGVHLINSCTFQSQTEFQKIQYNIVP	463
sp B0R7U1 DP2S_HALS3	GRTRLAPEAEDHLAIDTVPDFVHTGHVHKLGVGIHHNVRIVNSGCWQHQTAQFQESVNISP	487
sp O28484 DP2S_ARCFU	GRTPLAPEREDYLVIEDVPDILHCGHIHTYGTGFYRGVFMVNSSTWQAQTEFQKKVNLNP	466
sp P81412 DP2S_PYRFU	GKVPiapdpedllvievvpdvvhmgvhvyydavvyrgvqlvnnsatwqaqtefqkmvnivp	588
sp Q58113 DP2S_METJA	GRCPIAPEHKDYLVIDRDIDILHTGHIIHNGYGIYRGVVMVNSCTFQEQTDFQKRMGISP	563
	. : .. * .. : .. :***: .. : .. * : *** : * : .. * : .. *	

Fig. 2B. CLUSTAL format for T-COFFEE version of Pol II polymerases (Prokaryotes)

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

Fig. 2C Prokaryotic DNA polymerases (Poli III)

tr Q8G7V8 Q8G7V8_BIFLO	FCGYFLVVADYI QWAKDHGIMV GPGRGSAGAM VAYSMGITELDPLKHGLIFERFLNPER	418
tr C4ZRS6 C4ZRS6_ECOBW	FPGYFLIVMEFI QWSKDNGPV PGPGRGGAGS LVA YALKIT DLDPLF DLL FERFLNPER	396
tr Q88MG5 Q88MG5_PSEPK	FPGYFLIVMDFI KWKAWNN DVPV PGPGRGGAGS LVA YVLKIT DLDPLA YD LLL FERFLNPER	398
tr C1DST1 C1DST1_AZOVD	FPGYFLIVMDFI KWKAWNN DVPV PGPGRGGAGS LVA YALKIT DLDPLA YD LLL FERFLNPER	399
tr Q1D042 Q1D042_MYXXD	FCGYFLIVQDFINWAKKM NIPV PGPGRGGAGS LVA YALRIT DVPD IPI YD LLL FERFLNPER	398
tr Q72ZD3 Q72ZD3_BACC1	FSDYFLIVWDFM KYAHEN HILT PGPGRGSAGS LVS YVLE ITDIDP EYD LLL FERFLNPER	378
tr Q97RC7 Q97RC7_STRPN	FDDYFLIVWDFL RFGQSNGYYM GMGRGSAGS LVS YVLA DITG IDPVEK NLL FERFLNPER	344
tr Q4QMF6 Q4QMF6_HAEI8	FPGYFLIVMEFI QWSKDNDI PV PGPGRGSAGS LVA YALKIT DLDPLF DLL FERFLNPER	396
tr Q2FG35 Q2FG35_STAA3	FEDYFLIVSDLI HYAKTN DVMV PGPGRGSAGS LVS YLLGITTIDP IKF NLL FERFLNPER	342
tr A1R5G9 A1R5G9_ARTAT	FPGYFLIVVADFINWAKNN GIRV PGPGRGSAGS MVA YAMRIT DLDPLQ HGLIFERFLNPDR	419
tr Q6NGE6 Q6NGE6_CORDI	YPSYFLIVAE LIKHAR SIGIRV PGPGRGSAGS ALV AYALTIT NI DPMEH GLL FERFLNPER	416
tr Q6D8C8 Q6D8C8_ERWCT	FPGYFLIVMEFI QWSKDNDV PV PGPGRGSAGS LVA YSLKIT DLDPLF DLL FERFLNPER	396
tr Q7NZL1 Q7NZL1_CHRVO	FPGYFLIVADFI QWGK GCPV PGPGRGSAGS LVA YSLSIT DLDPLK YALL FERFLNPER	398
tr Q5NM98 Q5NM98_ZYMMO	FPGYFLIVADFI WAKSNQ I PV PGPGRGSAGS SVVA WALTIT NLDP RL GLL FERFLNPER	347
tr Q6KHF7 Q6KHF7_MYCMO	FADYFLIIY DIVKFA KENKI SV PGPGRGSAGS SIVSY LLNITS INPL KYD LLL FERFLNPER	324
tr C3K6H1 C3K6H1_PSEFS	FPGYFLIVMDFI QWAKS NGP V PGPGRGSAGS LVA YVQ KIT DLDPL EYD LLL FERFLNPER	398
tr B8ZR78 B8ZR78_MYCLB	FPAYFLIVADLV NHAR SVG IRV PGPGRGSAGS SLA AYALG IT DIDI P I PH GLL FERFLNPER	412
tr A1F2E4 A1F2E4_VIBCH	FPGYFLIVMEFI QWSKDNDI PV PGPGRGSAGS LVA YALKIT DLDPL EYD LLL FERFLNPER	401
tr C7C104 C7C104_HELPB	FPGYMLIVWDFI RYAKEM GIPV PGPGRGSAGS LVA FALKIT DIDI PLK YD LLL FERFLNPER	420
tr Q2K9R0 Q2K9R0_RHIEC	FPGYFLIVSDFI KWKAH DIPV PGPGRGSAGS LVA YALTIT D VDPL RFS LLL FERFLNPER	423
tr B5Y1I7 B5Y1I7_KLEP3	FPGYFLIVMEFI QWSKDNDV PV PGPGRGSAGS LVA YALKIT DLDPLF DLL FERFLNPER	396
tr Q2KYR1 Q2KYR1_BORA1	FPGYFLIVQDFINWGNNGP V PGPGRGSAGS LVA YALG IT DLDPL IRYD LLL FERFLNPER	401
tr Q5HV63 Q5HV63_CAMJR	FSGYMLIVHDFI VAKD KGIPV PGPGRGSAGS LVS YCL RIT DLDPL I PYD LLL FERFLNPER	421
:	: * : * : : * : * * * . * : * : * : * : * :	

tr Q8G7V8 Q8G7V8_BIFLO	VSLPDIDVDFDPEGRARVIEYC GEKY GTDKVAQC V IYGT IKTQ ALKDSARIM-GY-EFS	476
tr C4ZRS6 C4ZRS6_ECOBW	VSPMPDFDVDFCMEKRDQVIEHVADM YGRDAV S QI IT FGTMAAKAVIRDVGRVL-GH-PYG	454
tr Q88MG5 Q88MG5_PSEPK	VSPMPDFDVDFCMDGRDRV IDYVAE YGRNAV S QI IT FGTMAAKAVVRDVARVQ-GK-SYG	456
tr C1DST1 C1DST1_AZOVD	ISMPDFDVDFCMDGRDRV IDYVA DTYGRNAV S QI IT FGSMAAKAVVRDVARVQ-GK-SYG	457
tr Q1D042 Q1D042_MYXXD	VSPMPDFDIDE CQDDR DEV I QYVGR KYGEM NVG QI IT FGSLKA KS VL RDVCRVF-AL-PFS	456
tr Q72ZD3 Q72ZD3_BACC1	VTLPDIDIDF PDIRDE MIR YVK DKY QGL RV A QIV T FGT LAAKA AIR DIAR VM-GL-PPR	436
tr Q97RC7 Q97RC7_STRPN	YTMPDIDIDF PDI YRP DFIR YV GNK YGSK HAAQ I VFT STFG AK QAL RDV LKRF-GV-PEY	402
tr Q4QMF6 Q4QMF6_HAEI8	VSPMPDFDVDFCMDGRDRV IEHVA ETYRG AV S QI IT FGTMAAKAVIRDVGRVL-GH-PYG	454
tr Q2FG35 Q2FG35_STAA3	VTPMPDIDIDF EDTR RER V I QYVQE KYGEL R VSG I VT FG HLLA RA VAR DV GRIM-GF-DEV	400
tr A1R5G9 A1R5G9_ARTAT	VSPMPDFDVDFDDRR RSE V IDYV T K KYG DER V AMIV T YGT IKTQ ALK DSS RVL-GY-PFS	477
tr Q6NGE6 Q6NGE6_CORDI	PSAPDIDIDF DDR RRG EMIR YAA QRW GED KIA QV IT FGT VTKT Q ALK DSAR V QY QEG YK	476
tr Q6D8C8 Q6D8C8_ERWCT	VSPMPDFDVDFCMEKRD M V IDHVA E MYGRDAV S QI IT FGTMAAKAVIRDVGRVL-GH-PYG	454
tr Q7NZL1 Q7NZL1_CHRVO	VSPMPDFDVDFCQENRWRV IEYTR R KYG EEE AV S QI AT FGT MSSK S VIRD VGRVL-DL-PFG	456
tr Q5NM98 Q5NM98_ZYMMO	VSPMPDFDIDE CET RR SE V IS YVQ HKY G LDHVA QI IT FG RGM KAR VLA K DT G RVL-QM-SYG	405
tr Q6KHF7 Q6KHF7_MYCMO	ITMPDIDIDF QDDR RDE I VEYV T K KYG HQ K VAL IT FQR FGA K M AL RDV GRNL-NF-TQV	382
tr C3K6H1 C3K6H1_PSEFS	VSPMPDFDVDFCMDGRDRV IEYVA E KYG RNAV S QI IT FG S MAAKAVIRDVARVQ-GK-SYG	456
tr B8ZR78 B8ZR78_MYCLB	TSPMPDIDIDF DDR RRG EM V RY A ADK WGH DRV A QV IT FGT IKT K A ALK DSARIHYG QPG FA	472
tr A1F2E4 A1F2E4_VIBCH	VSPMPDFDVDFCMDKRDQV IDHVA E MYGRDAV S QI IT FGTMAAKAVIRDVGRVL-GH-PFG	459
tr C7C104 C7C104_HELPB	ISMPDIDIT DFC QRR RKE I IE YMIE K YG KYN V A QV IT FN KMLA K G VIRD VAR VL-DM-PYK	478
tr Q2K9R0 Q2K9R0_RHIEC	VSPMPDFDIDE CQDR REEV IR YV QAK YGRE QVA QI IT FG S L QARA AL RDV GRVL-EM-PYG	481
tr B5Y1I7 B5Y1I7_KLEP3	VSPMPDFDVDFCMEKRDQVIEHVADM YGRDAV S QI IT FGTMAAKAVIRDVGRVL-GH-PYG	454
tr Q2KYR1 Q2KYR1_BORA1	VSPMPDFDIDE CQDN RER V IEV KLY G RAA V S QI AT FG TLG A K VAV RDAGR VL-DM-PYM	459
tr Q5HV63 Q5HV63_CAMJR	VSPMPDIDVDFCQDRRAE VIDYV IDK YGAD KVA QV IT FG KLLA K G VIRD VAR VC-DM-SIQ	479
:	: * : * : * : * * : : : * : * :	

tr Q8G7V8 Q8G7V8_BIFLO	TLGLVKMDFLGI	SNLTIIHDTLKNIEANG-----K-PAIDYTKIPLDDAETYKLMSR	630
tr C4ZRS6 C4ZRS6_ECOBW	YAGLVKFDFLGI	RTLTIIINWALEMINRKRAK---NGE-PPLDIAAIPLDDKSFMLQR	602
tr Q88MG5 Q88MG5_PSEPK	AAGLVKFDFLGI	RTLTIIKWAMEIINREQAK---KNL-PDLNIDFIPDDRKTLYELLQK	604
tr C1DST1 C1DST1_AZOVD	AAGLVKFDFLGI	RTLTIIKWAMETIHREQR----RGETELVDIDRIALDDKATYALLQR	606
tr Q1D042 Q1D042_MYXXD	AAGLVKFDFLGI	KTLLTVIQLHALDVKRNH-----G-KDIPRHIPLNDEKTwELMAK	610
tr Q72ZD3 Q72ZD3_BACC1	EIGLLKMDFLGI	RNLTLLENIIKFIVQKT-----G-KEIDIRNLPLQDEKTFQLLGR	579
tr Q97RC7 Q97RC7_STRPN	ASGLLKMDFLGI	RNLTFVQKMQUELLA-----G-IHLKIEEIDLEDKETLALFAS	542
tr Q4QMF6 Q4QMF6_HAEI8	YAGLVKFDFLGI	RTLTIIKWALDIINVRMVR---EGK-PRVDIAAIPLDDPESFELLKR	602
tr Q2FG35 Q2FG35_STAA3	RIGLLKIDFLGI	RNLSIIHQILTVQVKKDLGI-----NIDIEKIPFDDQKVFELOSSQ	541
tr A1R5G9 A1R5G9_ARTAT	GLGLIKMDFLGI	RNLTIIISDALENIKMNR-----G-VDLDLENLELDDAASYELLAR	631
tr Q6NGE6 Q6NGE6_CORDI	AIGLLKMDFLGI	RNLTIVIGDCLENIKINR-----N-EELDLEALQTDDPNVYKLLSS	630
tr Q6D8C8 Q6D8C8_ERWCT	YAGLVKFDFLGI	RTLTIIIDWALEMINARRAK---LGQ-EPIDIATIPLGDKKSFDMLQR	602
tr Q7NZL1 Q7NZL1_CHRVO	QIGLVKFDFLGI	RNLTIIIELAQYIKDLT-----G-EDVDAVHLPLDDKPAYKVFAF	602
tr Q5NM98 Q5NM98_ZYMMO	KTGLVKMDFLGI	KTSLVNLNAQVKKLAR-----G-VEVDLHDIPLDDEAVFELLKQ	550
tr Q6KHF7 Q6KHF7_MYCMO	NFSLLKIDFLGI	KLASTIKLIEQKIREAYKIKYNFSNDEYEYTEQLKKDLPKKAFQLLSD	531
tr C3K6H1 C3K6H1_PSEFS	AAGLVKFDFLGI	RTLTIIDWALKTINRDRAK---VNE-EPLDIAFIPLDDKPTYSLLQK	604
tr B8ZR78 B8ZR78_MYCLB	AIGLLKMDFLGI	RNLTIIIGDAIENIKTNR-----G-IDLDLESVPPLDDQATYELLGR	626
tr A1F2E4 A1F2E4_VIBCH	TAGLVKFDFLGI	RTLTIIIDWALGLVNPRLKK---AGK-PPVRIEAIPPLDDARSFRNLQD	607
tr C7C104 C7C104_HELPB	PVDLIKFDLGI	KTLLTVIIDDALKITKTOH-----N-IDVDFLISLMDDPKVYKTIQS	633
tr Q2K9R0 Q2K9R0_RHIEC	QAGLVKFDFLGI	KTLLTVKAVDFVAKR-----G-INVDLAAIPLDDKPTYEMLSR	626
tr B5Y1I7 B5Y1I7_KLEP3	YAGLVKFDFLGI	RNLTIIINWALEMINRKRAK---NGE-GPLDIAAIPLDDKKSFDMLQR	602
tr Q2KYR1 Q2KYR1_BORA1	AAGLVKFDFLGI	RNLTILDWAVRYVRQFN-----PDQ-RDFDIMALSLDDPAAYKVLCD	608
tr Q5HV63 Q5HV63_CAMJR	DVDLIKFDLGI	KTLLTIVINNAIKLIKRY-----N-KDIIWETIDVNDSKVYKTIQS	625

. * : * : * : * : * : : : :

tr Q8G7V8 Q8G7V8_BIFLO	GDTLGVFQLDSDGMRSLLKTLKPDNFNDISALIALYRPGFMSMDSHTNYAKRKNGLQKIT	690
tr C4ZRS6 C4ZRS6_ECOBW	SETTAVFQLESRGMKDLIKRLQPDCFEDMIALVALFRPGFLQSGMVDFNIDRKHGREEIS	662
tr Q88MG5 Q88MG5_PSEPK	AETTAVFQLESRGMKELIKKLKPDCLEDLIALVALFRPGFLQSGMVDFINRKHGRAELA	664
tr C1DST1 C1DST1_AZOVD	AETTAVFQLESRGMKELIKKLKPDNIEDMIALVALFRPGFLQSGMVDFINRKHGRAELS	666
tr Q1D042 Q1D042_MYXXD	GDTAGIFQMESSGFTEMVVKLKPNCFEDVIAAGALYRPGFLDSGMVDVFIRKHGREKVS	670
tr Q72ZD3 Q72ZD3_BACC1	GDTTGVFQLESGGMRNVLRLQPKNEFEDIVAVNSLYRPGFME--QIPTFIESKHGKRKIE	637
tr Q97RC7 Q97RC7_STRPN	GNTKGIFQFEEQPGAIRLLKRVQPCFEDVVATTSLNRPGASD--YINNFVARKHGQEEVT	600
tr Q4QMF6 Q4QMF6_HAEI8	SETTAVFQLESRGMKDLIKRLQPDCFEDIIALVALFRPGFLQSGMVDFNIDRKHGREEVS	662
tr Q2FG35 Q2FG35_STAA3	GDTTGIFQLESQDGVRSVLLKLPKPEHFEDIVAVTSLYRPGFME--IPTYITRRHDPSKVQ	599
tr A1R5G9 A1R5G9_ARTAT	GDTLGVFQLDGGPMRSLLKLMKPDPNFEDISAVLALYRPGPMGANHTDYALRKNGIQBEVI	691
tr Q6NGE6 Q6NGE6_CORDI	GDTLGVFQLDGGMQUELLKRMQPTGFNDIVASLALYRPGPMGVNAHDYADRKNRKPPIE	690
tr Q6D8C8 Q6D8C8_ERWCT	SETTAVFQLESRGMKDLIKRLKPDCFEDMIALVALFRPGFLQSGMVDFNIDRKHGREAIS	662
tr Q7NZL1 Q7NZL1_CHRVO	ANITAVFQFESTGMKKMLEAKPSKFEIIIAFVALYRPGFMD--LIPDFIQRMHGA-KFE	659
tr Q5NM98 Q5NM98_ZYMMO	GDTVGVFQLESEGMRRTLAAPKTRFEDIIALVALYRPGFMDN--IPSFGARKNGEEIV	608
tr Q6KHF7 Q6KHF7_MYCMO	GNTAGVFQLESQPMIQTICKVSKISKFSDIVDVISLYRPGFMQY--IQEYIDNKNDASKIK	589
tr C3K6H1 C3K6H1_PSEFS	AETTAVFQLESRGMKELIKKLKPDCLEDLIALVALFRPGFLQSGMVDDFINRKHGRAELA	664
tr B8ZR78 B8ZR78_MYCLB	GDTLGVFQLDGGPMRDLLRMRQPTGFEDIVAVLALYRPGPMGMNAHDYADRKNRQVIK	686
tr A1F2E4 A1F2E4_VIBCH	AKTTAVFQLESRGMKELIKRLQPDCFEDIIALVALFRPGFLQSGMVDFNIDRKHGREAIS	667
tr C7C104 C7C104_HELPB	GDTVGIFQIESGMFQGLNKLRLPSSFEDIIIAIILGRPGPMESGMVDDFVNRKHGVEPIA	693
tr Q2K9R0 Q2K9R0_RHIEC	GETVGVFQVESAGMRKALIGMKPDCIEDIIALVALYRPGPMEN--IPVYNARKHGEEEEE	684
tr B5Y1I7 B5Y1I7_KLEP3	SETTAVFQLESRGMKDLIKRLQPDCFEDMIALVALFRPGFLQSGMVDFNIDRKHGREEIS	662
tr Q2KYR1 Q2KYR1_BORA1	ANTTAVFQLESRGMKELLKKLRPNTFEDIIAMLALYRPGPLESGMVDDFVNRKHGRAAVD	668
tr Q5HV63 Q5HV63_CAMJR	GNTLGIFQIESGGMQSLNARLKPERFEDIIAVLALYRPGPMESGMLDDFIDRKHGGLKNE	685

.. * . : * . : : : : * *** : :

tr Q8G7V8 Q8G7V8_BIFLO	PIHPEL-EEPLKQVLDETYGLIILYQEQQVOSAARILAGYSLGKADVLRRAMGKKPVELAK	749													
tr C4ZRS6 C4ZRS6_ECOBW	YPDVQWQHESLKPVLEPTYGILILYQEQQVQIAQVLSGYTLGGADMRLRAMGKKPEEMAK	722													
tr Q88MG5 Q88MG5_PSEPK	YPHSDYQYEGLKPVLAPTYGILILYQEQQVQIAQVMSAGYTLGGADMRLRAMGKKPEEMAK	724													
tr C1DST1 C1DST1_AZOVD	YPHPDYQYAGLEPVLPKPTYGILILYQEQQVQIAQVMSAGYTLGGADMRLRAMGKKPEEMAK	726													
tr Q1D042 Q1D042_MYXXD	YPHPA----LEPVLPKDTYGVIVYQEQQVQIISQVLGGYTLGRADLRLRAMGKKKAEVMQA	725													
tr Q72ZD3 Q72ZD3_BACC1	YLH----PDLKPLERTYGVIVYQEQQVQIAASKLAGFSLGEADLLRRAVSKKNRDILDQ	692													
tr Q97RC7 Q97RC7_STRPN	VLD----PVLEDILAPTYGILILYQEQQVQVAQRLAGFSLGKADILRRAMGKKDASAMHE	655													
tr Q4QMF6 Q4QMF6_HAEI8	YPDAEYQHASLKPILEPTYGILILYQEQQVQIAQVLAGYTLGGADLRLRAMGKKPEEMAK	722													
tr Q2FG35 Q2FG35_STAA3	YLH----PHLEPILKNTYGVILYQEQQIMQIASTFANFSYGEADILRRAMSKKNRAVLES	654													
tr A1R5G9 A1R5G9_ARTAT	PIHPEL-EEPPLKEILGGTFGLIVYQEQQMAVAQKLAGYSLGQADILRRAMGKKKSELDK	750													
tr Q6NGE6 Q6NGE6_CORDI	PIHPEL-EEPPLKEILEETYGLIVYQEQQIMEISRKLANYTAGEADGFRKAMGKKPVELAK	749													
tr Q6D8C8 Q6D8C8_ERWCT	YPDIEWQHESLKPVLEPTYGILILYQEQQVQIAQVLAGYTLGGADMRLRAMGKKNPVEMAK	722													
tr Q7NZL1 Q7NZL1_CHRVO	YLH----PILLEPVLPAPTYGILIVYQEQQVQMAQVQVLAGYSLGGADLRLRAMGKKVEMVA	714													
tr Q5NM98 Q5NM98_ZYMMO	YPH----PILLEPVLKETTYGILIVYQEQQVQAAQVLAGYSLGEADLRLRAMGKKIQSEMDA	663													
tr Q6KHF7 Q6KHF7_MYCMO	HVE----KLYDEIVSKTHGILILYQEQQILEIVQKYAGLSYAKADILRRRAISKNNSEEIEK	644													
tr C3K6H1 C3K6H1_PSEFS	YPHSDYQYEGLKPVLAPTYGILILYQEQQVQIAQVMSAGYTLGGADMRLRAMGKKPEEMAK	724													
tr B8ZR78 B8ZR78_MYCLB	PIHPEL-EEPPLREILAETYGLIVYQEQQIMRIAQKVAGYSLARADILRKAMGKKREVLEK	745													
tr A1F2E4 A1F2E4_VIBCH	YPDEKWQHESLKPILEPTYGILILYQEQQVQIAQVLSGYTLGGADMRLRAMGKKPEEMAK	727													
tr C7C104 C7C104_HELPB	YAF----KELEPILKPTYGILIVYQEQQVQIIVQTIGGFSLGEADLRLRAMGKKDAQIMAD	748													
tr Q2K9R0 Q2K9R0_RHIEC	SIH----PTIDHLLKETQGVIVYQEQQVQIAQVLSGYSLGEADLRLRAMGKKIKAEQMDQ	739													
tr B5Y1I7 B5Y1I7_KLEP3	YPDVQWQHESLKPVLEPTYGILILYQEQQVQIAQVLSGYTLGGADMRLRAMGKKPEEMAK	722													
tr Q2KYR1 Q2KYR1_BORA1	YFH----PDLEGTLKSTYGVIVYQEQQMLISQIIGGYSLGGAIDLRRAMGKKPEEMAK	723													
tr Q5HV63 Q5HV63_CAMJR	YPF----DSLEKVLEPTYGVIVYQEQQVQIVQIIGGFSLGGADVVRAMGKKDPEKMKK	740													
	:	*	*	:	*****:	..	:	..	**	.*	:*	..	**	:	..

tr Q8G7V8 Q8G7V8_BIFLO	EKVPFFAGMKEHGYSSEASQAVWDILVPFGSGYAFNKAHSAAAYGILISYWTAYLKTHYPVEF	809														
tr C4ZRS6 C4ZRS6_ECOBW	QRSVFAEGAENGNINAELAMKIFDLVFKAGYGFNKSNSAAYAIVSYQTLWLKAHYPAEF	782														
tr Q88MG5 Q88MG5_PSEPK	QRGGFIEGCVANNIDADLAGNIFDLVFKAGYGFNKSNSAAYGIVSYQTLWLKTHYPAPF	784														
tr C1DST1 C1DST1_AZOVD	QRGGFVEGCAKNGIDAEALAGNIFDLVFKAGYGFNKSNSAAYGIVSYQTLWLKTRYPAPF	786														
tr Q1D042 Q1D042_MYXXD	ERAGFLEGCAKNNVDLKVAGEIFDLMEKFAEYGFNKSNSAAYGIVTHTAWLKAHYPCEF	785														
tr Q72ZD3 Q72ZD3_BACC1	ERKHFVQGCLQNQYDETSAEKIYDLIVRFANYGFNRSHAVAYSNIIGYQLAYLKANYTLEF	752														
tr Q97RC7 Q97RC7_STRPN	MRASFIQGSLEAGHTVEKAEEQVFDVMEKFAGYGFNKSNSAAYGIVSYQTLWLKTHYPAIF	715														
tr Q4QMF6 Q4QMF6_HAEI8	QRLVFKEGAEKNGVNDGELSLSMKIFDLVFKAGYGFNKSNSAAYAIVSYQTLWLKTHFPAEF	782														
tr Q2FG35 Q2FG35_STAA3	ERQHFIEGAKQNGYHEDISKQIFDLILKFADYGFPRAHAVSYSKIAIMSFLKVHYPNYF	714														
tr A1R5G9 A1R5G9_ARTAT	QFAGFSQGMQDNGYSMAAVKTLWDLILLPFSDYAFNKAHSAAAGYIVSYWTAYLKAHYAPEY	810														
tr Q6NGE6 Q6NGE6_CORDI	EYKKFSEGMSNGYSKAAVDALWGTIEPFASYAFNKSNSAAGYIVSYWTAYLKANYTAEY	809														
tr Q6D8C8 Q6D8C8_ERWCT	QRGGFEDGAKSRGVNGELAVKIFDLVFKAGYGFNKSNSAAYAIVSYQTLWLKAHYPAEF	782														
tr Q7NZL1 Q7NZL1_CHRVO	QRAMFVQGAQKDIPAKEANEIFDYMFKAGYGFNKSNSAAYAIVYHTAWLKAHHCAAY	774														
tr Q5NM98 Q5NM98_ZYMMO	QRSRFVEGSKKHIDTENQANQLFDLIDFKAGYGFNKSNSAAYAIVYQTAWLKAHYRAEF	723														
tr Q6KHF7 Q6KHF7_MYCMO	FKDLFISSSIQHGQSEKVNANEIFSKEIQAQYGFNKSNSAAYAIVSYQTLWLKANFPLEF	704														
tr C3K6H1 C3K6H1_PSEFS	QRGGFIEGCATNNIDADLAGNIFDLVFKAGYGFNKSNSAAYGIVSYQTLWLKAHYPAEF	784														
tr B8ZR78 B8ZR78_MYCLB	EFEFGFSEGQMANGFSVHAIKALWDIILPFDADYAFNKSNSAAGYILISYWTAYLKANFAGEY	805														
tr A1F2E4 A1F2E4_VIBCH	QRAVFQEGAEKNGVNDGELAMKIFDLVFKAGYGFNKSNSAAYAIVSYQTLWLKTHYPVEF	787														
tr C7C104 C7C104_HELPB	NKAKFVEGAKNLGHGDQKAANLWDLIVFKAGYGFNKSNSAAYAINTFQTAWLKTYYKHEF	808														
tr Q2K9R0 Q2K9R0_RHIEC	QRERFVEGAKVNGVKKPQADVIFDLLAKFANYGFNKSNSAAYAIVSYQTAWMKAHYPVEF	799														
tr B5Y1I7 B5Y1I7_KLEP3	QRSIFEDGAKKNGVNDGELAMKIFDLVFKAGYGFNKSNSAAYAIVSYQTLWLKAHYPAEF	782														
tr Q2KYR1 Q2KYR1_BORA1	HRELFQKGAVDKHDPDLAVKLFDLMEKFAGYGFNKSNSAAYAIVYQTAWLKAYHPAEF	783														
tr Q5HV63 Q5HV63_CAMJR	LKTDFADGAEKQGYDRVKAEDLWEILVKFAGYGFNKSNSAAYAINTFQTAWLKTYYPSEF	800														
	*	.	.	:	:	*	.	*	:	*	*	:	*	:	*	.

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-----RKEIHVDMDCFFAAVEMRDPRLRDIPLAIGGSKE---RRGVISTANY	46
sp Q99Y66 DPO4_STRP1	MLIFP-LINDTSRKIIHIDMDAFFAAVEERDNPALKGKPVIGKDPRETGGRVVSTCN	59
sp P63990 DPO4_SALTI	M-----RKEIHVDMDCFFAAVEMRDPALRDPIAIGGSRE---RRGVISTANY	46
sp P58964 DPO4_STAAW	M-----TERRIIHIDMDYFFAQEMERDNPALKGKPVIVGGKAS---SRGVVSTASY	48
sp P63989 DPO4_SALTY	M-----RKEIHVDMDCFFAAVEMRDPALRDPIAIGGSRE---RRGVISTANY	46
sp Q5M2Y5 DPO4_STRT2	MLEFP-LINDTSRKIIHIDMDAFFAQEMERDDPSLKDGPVIIGNDRKTGGRVVSTCN	59
sp Q4ZWM4 DPO4_PSEU2	MT-----QRKEIHIIDCDCFYAAIEMRDEPDLAGKPLAVGGSAE---RRGVIACTCN	48
sp A7Z6F3 DPO4_BACA2	MAG-----KGRIEFHIDMNSFYASVEMAYDPSLRGKPIAVAGNVKE--RKGIIVTCSY	51
sp Q9AK82 DPO4_STRCO	M-----RTAPTLHLMDMAFFAVERQASKPSLRGKAVVVGGLG---PRGVVATCSY	48
sp C3LQ59 DPO4_VIBCM	MQ-----DIRKKEIHVDMDCFFAAVEMRDPALRDPIAIGGSRE---RRGVISTANY	50
sp Q8X7Q1 DPO4_ECO57	M-----RKEIHVDMDCFFAAVEMRDPALRDPIAIGGSRE---RRGVISTANY	46
sp Q83M86 DPO4_SHIFL	M-----RKEIHVDMDCFFAAVEMRDPALRDPIAIGGSRE---RRGVISTANY	46
sp Q8XK37 DPO4_CLOPE	M-----KENRKIIHIDMDAFFAVERQDRDNPKYKGKPLIVGGDPN---RRGVVATCSY	49
sp Q635E0 DPO4_BACZ	MSTMREMPYKKGRVILHVDMNCFFAAVEMRDPALRDPIAIGGSRE--RKGIIVTCSY	58
sp Q88V07 DPO4_LACPL	M-----DTSRKIIHVDMDCFFAAVEMRDPALRDPIAIGGSRE--RKGIIVTCSY	52
sp Q8NNP4 DPO4_CORGL	M-----QRWLHIDMDAFFAVERQDRPQLTRPRLGRPVLVGGVS---GRGVVAGASY	46
sp Q9KPS5 DPO4_VIBCH	MQ-----DRIRKKEIHVDMDCFFAAVEMRDPALRDPIAIGGSRE---RRGVISTANY	50
sp Q6A763 DPO4_PROAC	M-----RSTASLHLMDMAFFAVERQDRDNPKYKGKPLIVGGVG---GRGVVATASY	48
sp Q88NK4 DPO4_PSEPK	MS-----LRKKEIHVDCDCFYAAIEMRDDPRLAGRPMAVGGSPD---HRGVIACTCN	48
sp Q8DVR7 DPO4_STRMU	MLIFP-LINDTSRKIIHIDMDAFFAAVEERDNPALKGKPLIVGADPRQTGGRVVSTANY	59
*	: * : * : * : * : . : . : : : * : * : . : * : : * : .. *	

sp Q8ZBZ9 DPO4_YERPE	PARRYGVRSAMPTAMAFK-LCPQLTLLPGRMAAYKEASQHIREIFARYTPLIEPLSLDEA	105
sp Q99Y66 DPO4_STRP1	EARKYGIHSAMSSKEAYE-RCPKAIFISGNYEKYRTVGDQIRRIFKRYTDVVEPMIDEA	118
sp P63990 DPO4_SALTI	PARQFGVRSAMPTAMALK-LCPHLLPGRFDAYKEASRHVRDIFSRYTSLIEPLSLDEA	105
sp P58964 DPO4_STAAW	EARKYGVHSAMPMSQAHK-LCPNGYFTSNFGAYRETSQAQIMSIFFRSYTDKVEPMIDEA	107
sp P63989 DPO4_SALTY	PARQFGVRSAMPTAMALK-LCPHLLPGRFDAYKEASRHVRDIFSRYTSLIEPLSLDEA	105
sp Q5M2Y5 DPO4_STRT2	EARKYGVHSAMSSKEAYE-RCPNAVFISGNYSHREVMQIREFKCYTDLVEPMIDEA	118
sp Q4ZWM4 DPO4_PSEU2	EARAYGVRSAMSSRHALK-LCPDLTIVKPRMDAYKEASREIHTIFRDTDLIEPLSLDEA	107
sp A7Z6F3 DPO4_BACA2	EARARGVTTMPVWRDKR-LCPPELIVVPPNFDRYRSSSKEMFSILREYTDLVEPVSIDE	110
sp Q9AK82 DPO4_STRCO	EARVFGVHSAMPMQARR-LAPHAAYLVPRFELYRSISEQVMRLRELSPVLEPLSLDEA	107
sp C3LQ59 DPO4_VIBCM	QARKFGVRSAMPTAQALK-LCPQLHVVPGRMSVYKSVQQIQTIFQRTSLSIEPLSLDEA	109
sp Q8X7Q1 DPO4_ECO57	PARKFGVRSAMPTGMALK-LCPHLLPGRFDAYKEASNHIREIFSRTSRIEPLSLDEA	105
sp Q83M86 DPO4_SHIFL	PARKFGVRSAMPTGMALK-LCPHLLPGRFDAYKEASNHIREIFSRTSRIEPLSLDEA	105
sp Q8XK37 DPO4_CLOPE	EARKYGIHSAMPSLTAYK-LCPKAIFIRPMEVYKKVSRQVMNLNEYNSNIVEPLSLDEA	108
sp Q635E0 DPO4_BACZ	EAREYGIRTTMPWLREAKR-LCPQLIVRPNFTLYREASFQMFQILSRFTEKIQPVSIDE	117
sp Q88V07 DPO4_LACPL	VARQFGVHSAMPAKALE-LCPTAVFKTPNFPFLYREVSQAQIHRIFHEYTEMIEPIAFDEA	111
sp Q8NNP4 DPO4_CORGL	EARKFGARSAMPMHQAKARVGFAGVVTPRHIVYSAASRRVFQIVEKRAGIVERLSIDE	106
sp Q9KPS5 DPO4_VIBCH	QARKFGVRSAMPTAQALK-LCPQLHVVPGRMSVYKSVQQIQTIFQRTSLSIEPLSLDEA	109
sp Q6A763 DPO4_PROAC	EARKFGVHSAMAGSQARR-LAPNAFLSGRFESYRESSKVMATLRELSPVLEPLSLDEA	107
sp Q88NK4 DPO4_PSEPK	EARAYGVRSAMSSRHALK-LCPDLTIVKPRFEAYREASREIHTIFRDTIELIEPLSLDEA	107
sp Q8DVR7 DPO4_STRMU	LAREFGIHSAMSSKEAYE-RCPQAIFIRGNHTKYRQIGLQVREIFRRTYTDLVEPMIDEA	118
*** * : : * . * . * . : . : : : : : : * : : : **		

sp Q8ZBZ9 DPO4_YERPE	YLDVSDSLAC---GGSATLIAQEIRQSIASE-LNLATASAGIAPIKFLAKIASELNKP---	159
sp Q99Y66 DPO4_STRP1	YLDVTDNKLG---IKSAVKIAKLIQHDIWKE-VGLTCAGVSYNKFLAKLASDFEKPH---	172
sp P63990 DPO4_SALTI	WLDVTDSPHC---YGSATLIAIREIRQTIFNE-LQLTASAGVAPVKFLAKIASDLNKP---	159
sp P58964 DPO4_STAAW	YLDITELVRP---DLPASKIAQYIRKDILEQ-THLTASAGVSYNKFIAKIASDLNKP---	161
sp P63989 DPO4_SALTY	WLDVTDSPHC---YGSATLIAIREIRQTIFNE-LQLTASAGVAPVKFLAKIASDLNKP---	159
sp Q5M2Y5 DPO4_STRT2	YLDVTTNKLG---IKSAVKVAKLIQYDIWQE-LHLTCAGVSYNKFLAKLASDFQKPA---	172
sp Q4ZWM4 DPO4_PSEU2	FLDVSEAGHF---SGSATRIAQDIRRRVSNQ-LHITVSAGVAPPNKFLAKIASDWKKNP---	161
sp A7Z6F3 DPO4_BACA2	YMDITDTPH---SHRAYETAEDIQARLQKE-LLLPPSSIGIAPPNKFLAKMASDMKKP---	163
sp Q9AK82 DPO4_STRCO	FVLDLADGAA-RDAETARLAGTKLRTDIRTV-TGLTGSGVLAASKMLAKIASEAAKPD---	163
sp C3LQ59 DPO4_VIBCM	YLDVSESTAY---QGSATLIAQAIRDIWQE-LNLATASAGVAPIKFLAKVASDLNKP---	163
sp Q8X7Q1 DPO4_ECO57	YLDVTDHSVHC---HGSATLIAQEIRQTIFNE-LQLTASAGVTPVKFLAKIASDMNKP---	159
sp Q83M86 DPO4_SHIFL	YLDVTDHSVHC---HGSATLIAQEIRQTIFNE-LQLTASAGVAPVKFLAKIASDMNKP---	159

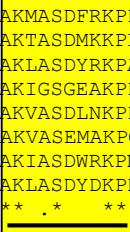
<pre> sp Q8XK37 DPO4_CLOPE FVDVSKSKRC---KGSATLIALEIKERIFKE-VGLTASAGVSFNKFL<u>AKMASDFRKPD</u>-- 162 sp Q635E0 DPO4_BACZ YLDITDCYA---LGSPLEIAKMIQQALLTE-LQLPCSIGIAPNLFL<u>AKTASDMKKPL</u>-- 170 sp Q88V07 DPO4_LACPL YLDVTNKKH---IHSAVELAHRLQQEIWHQ-THLTCSTGISYNKFI<u>AKLASDYRKPA</u>-- 165 sp Q8NNP4 DPO4_CORGL FMEPEALVGA--TPEEVKQWAEELRAEIKEV-TGLPSSVGAGSGKQI<u>AKIGSGEAKPD</u>-- 161 sp Q9KPS5 DPO4_VIBCH YLDVSESTY---QGSATLIAQAIRRDIWQE-LNLTASAGVAPIKFL<u>AKVASDLNKPD</u>-- 163 sp Q6A763 DPO4_PROAC FVDLEAGDVDEVNLDALQQAVELRARVAERT<u>GLSCSVGIGSSKFM</u><u>AKVASEMAKPQRG</u> 167 sp Q88NK4 DPO4_PSEPK YLDVSDSQWY---SGSATRIEADIRRRVART-LHITVSGAVPNKFL<u>AKIASDWRKPN</u>-- 161 sp Q8DVR7 DPO4_STRMU YLDVTENKLN---IKSAVKIAKLI<u>QORDIWE</u>E-FHLTCAGSYNKFL<u>AKLASDYDKPH</u>-- 172 ::: . : : . : * * : ** . * **</pre>	<pre> ::: . : : . : * * : ** . * **</pre>
	
<pre> sp Q8ZBZ9 DPO4_YERPE ---GQYVITPNQIQPFLQDPLSKIPGVGA<u>VATAKRLQALGLVTCGDIQKYPLAELLKHFG</u> 216 sp Q99Y66 DPO4_STRP1 ---GLT^LVLKEDALCFLAKLPIEKFH<u>CGVKKSVKKLHDMDGIYTGQDILAVPEMTLIDHFG</u> 229 sp P63990 DPO4_SALTI ---GQYVITPADVPDFLKLTPLA<u>KIPGVGVKVA</u>AKLENMGLRTCGDIQQCDLAMLLKRGF 216 sp P58964 DPO4_STAAW ---GMTVIDYQNVDHILM^TL<u>DIGDFPGVGVKASKKVMHDNGI</u>FNGRDLYEKTEFELIRLFG 218 sp P63989 DPO4_SALTY ---GQYVITPADVPDFLKLTPLA<u>KIPGVGVKVA</u>AKLENMGLRTCGDIQQCDLAMLLKRGF 216 sp Q5M2Y5 DPO4_STRT2 ---GLTVVLP<u>PEAQEFLKKLPIAKFHV</u>GKVKSVERLHMDM^IYTGADILKISEITLIDRFG 229 sp Q4ZWM4 DPO4_PSEU2 ---GLFVITPDQVEEFVASL<u>PVTKLHG</u>VGVKTADKLGR^LGIVDCADLRSRSK^LALVREFG 218 sp A7Z6F3 DPO4_BACA2 ---GITILRKREL<u>PDVLPWPL</u>PIEEMH<u>GITKTAEKIKT</u>LG^IKTIGDIAGDEHALKTLLG 220 sp Q9AK82 DPO4_STRCO ---GLVLIPPGTERAMLEPM<u>PTVRTLPGVGPATGDH</u>L^RAGITTVGEIAEAGEDELVRLLG 220 sp C3LQ59 DPO4_VIBCM ---GLYVVTPDVKQEMVDSL<u>PLEKIPGVGVKVA</u>AKLEAMGLRTCGDVQKCDLVILLKRGF 220 sp Q8X7Q1 DPO4_ECO57 ---GQFVITPAEVA<u>SAFLQT</u>PLAKIPGVGVKVA<u>AKLEAMGLRTCGDVQKCDLVILLKRGF</u> 216 sp Q83M86 DPO4_SHIFL ---GQFVITPAEVA<u>PAFLQT</u>PLAKIPGVGVKVA<u>AKLEAMGLRTCGDVQKCDLVILLKRGF</u> 216 sp Q8XK37 DPO4_CLOPE ---GITVITEENSKDFIRKLP<u>IGKFFCVG</u>RVTKNKLNNIGVFKGEDILGFSEKELIGILG 219 sp Q635E0 DPO4_BACZ ---GITVLRKRDIPEM^IWPLPVGAMH<u>GIGE</u>KTAEKLN<u>DIHQ</u>TIQ^LQ^AKGNEHIIRAKIG 227 sp Q88V07 DPO4_LACPL ---GFTVLPQD<u>FLLREP</u>IEKFRG<u>VGKKTVPKMHD</u>LG^IKTQDLYAQSEL^DL^IKQFG 222 sp Q8NNP4 DPO4_CORGL ---G^FVVPVDP<u>DKQH</u>DL<u>DPLPVGALW</u>VG^GPV<u>TGSK</u>LSMGVETIGD^LALTQ^KVEI^SLG 218 sp Q9KPS5 DPO4_VIBCH ---GLYVVTPDVKQEMVDSL<u>PLEKIPGVGVKVA</u>AKLE^IQAGLYVGADVRADYRKLHQFG 220 sp Q6A763 DPO4_PROAC HTDQIALIAPGTEADTIA<u>PLSARA</u>IPGVGPVTAERLEK^LGLRT^IADVRAARESELVHELG 227 sp Q88NK4 DPO4_PSEPK ---GLFVITPNEVET<u>TFVA</u>ALPVARLHG<u>VGVK</u>TADKL^LTR^GIETCLH^IREWSRLALVREFG 218 sp Q8DVR7 DPO4_STRMU ---GLTVILPQDAEGFLATLPIEKFY<u>GVGVKK</u>SVEKLHALHIFTGKD^VQQVPEMTLIDLFG 229 ::: . : : . : * : . : . : : * :</pre>	<pre> ::: . : : . : * : . : . : : * :</pre>
	
<pre> sp Q8ZBZ9 DPO4_YERPE RGVR^LVG^IHVTLLDPQL--ERQLLL----- 349 sp Q99Y66 DPO4_STRP1 AG^IRLLG^VTMTNLEDKVA--DISL----- 362 sp P63990 DPO4_SALTI RGVR^LVG^IHVTLLDPQL--ERQLVL----- 349 sp P58964 DPO4_STAAW VP^IRLIGV^ITVGNLEQSTYK--NM^MTY----- 353 sp P63989 DPO4_SALTY RGVR^LVG^IHVTLLDPQL--ERQLVL----- 349 sp Q5M2Y5 DPO4_STRT2 SG^IRLLGV^ITLTGLEDQEGR--QLDLD----- 364 sp Q4ZWM4 DPO4_PSEU2 KPV^RLLGI^IGVRLHDLRAA-HEQLELF----- 351 sp A7Z6F3 DPO4_BACA2 NP^VRLLGI^ITGTTELVEKQQAYQ<u>KQLDLSFKEDAKDEPIF</u>----- 370 sp Q9AK82 DPO4_STRCO GGV^RLLGV^IGVSGLADY^TQE--DLFAQAAGDRA-EEPAAEPGTEP----- AEAHSP 378 sp C3LQ59 DPO4_VIBCM RE^IRLLG^ISVMLKPELQ--MKQLSMFP----- 355 sp Q8X7Q1 DPO4_ECO57 RGVR^LVG^IHVTLLDPQM--ERQLVL----- 349 sp Q83M86 DPO4_SHIFL RGVR^LVG^IHVTLLDPQM--ERQLVL----- 349 sp Q8XK37 DPO4_CLOPE KK^VRLLGI^ITISSLEENI^ITEREQLS----- 356 sp Q635E0 DPO4_BACZ DSV^RLLGV^ITATEIEWKTESVKQLDLSFEEADEAKEPLL----- 377 sp Q88V07 DPO4_LACPL SG^IRLLG^ITLTGLAPLAFE--NLTLPL----- 361 sp Q8NNP4 DPO4_CORGL GP^IRLGVG^ISFSGLEESRQD--ILFPELDQII-VPPAPD^TDYEVGVQSSSSSESTQVEAP 386 sp Q9KPS5 DPO4_VIBCH RE^IRLLG^ISVMLKPELQ--MKQLSMFP----- 355 sp Q6A763 DPO4_PROAC EG^VRLLGV^IGVSNFATSQAQE--ELFVV^DDEGRLLDEPEV---TEE----- ITPIQP 383 sp Q88NK4 DPO4_PSEPK KPV^RLLGV^IGVRLRDLRGA-HEQLELF----- 351 sp Q8DVR7 DPO4_STRMU SG^VRLLGV^ITVTALEDSTRE--ELSLTA----- 365 : * * : * :</pre>	

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)

gi 291617317 ref YP_003520059.1 gi 290792075 gb ADD63401.1 gi 52841464 ref YP_095263.1 gi 242376964 emb CAQ31686.1 gi 15672570 ref NP_266744.1 gi 6491876 gb AAF14056.1 AF039838_1 gi 292488580 ref YP_003531464.1 gi 150024494 ref YP_001295320.1 gi 291554011 emb CBA21065.1	-----MSLDTFGREIKTRIKKEVHLNVCVGIAPTKTLAKLANHAAKKWS-KTGGVL 163 -----MGLEDFGRQLRQHVYDCTRLTIGVGAGPTKTLAKSAQWASKEWK-QFRGVL 162 -----DSYDSFCCEQLQKKILKHTGIPTSIGIGPTKTLAKAANHLOKKVY-KI-PVF 166 -----RDLTDFGREIRATVLQRTHLTGVGIAQTKTLAKLANHAAKKWQRQTGGVV 164 FYPDIFRNRYEQMNLITLDLQREILDKLGLYVTVMGDNPPLAKLAMDNYAKHNQNMRALI 232 -----AKLANHAAKKWQEQTGGVV 19 -----VSLADFGHEMRNQVLRNTGLTVGVGIAQTKTLAKLANHAAKKWP-ATGGVV 163 -----YDLNDYCLTIRQVKVVGMPVSIGVAPTKAL\$KVANKIAKKFTDRTNGVY 162 -----VSLADFGHEMRNQVLRNTGLTVGVGIAQTKTLAKLANHAAKKWP-ATGGVV 163 * * : :
gi 291617317 ref YP_003520059.1 gi 290792075 gb ADD63401.1 gi 52841464 ref YP_095263.1 gi 242376964 emb CAQ31686.1 gi 15672570 ref NP_266744.1 gi 6491876 gb AAF14056.1 AF039838_1 gi 292488580 ref YP_003531464.1 gi 150024494 ref YP_001295320.1 gi 291554011 emb CBA21065.1	DLS--NRERQKKLLSIVPVA <u>D</u> IWGVGRRRLSQRLEEMGIKTADDLACQSTSIRKNFSVVL 221 ALTRGNPQRTRKLLSLQPVEEIWGVGNRIARKLNVLGIKTALDLALTNPAPFKRNFSVVL 222 NIT---SNRGRLQQISVGD <u>I</u> WGVGRQWANKLISRGIHTAYDLAMTNPHLLKKCFNAVL 222 DLS--NLERQRKLMSALPVDDVWGIGRRISKKLDAMGIKTVLDA <u>D</u> DIRFIRKHFNVL 222 R--YE--DVPSKLWTLPKMTDFWGIGKRTEKRLNKG <u>I</u> SSIKELTNADPLLKQKLGTIG 288 DLS--NIERQRKLMAALPVEDVWGIGRRIGKKLDVMGIKTVLDA <u>T</u> DIRFIRKHFNVL 77 DLT--HIDRQRKLLALVA <u>E</u> DVWGIGRRISKKLNLMGIETALDAESSLWVIRKHFNVL 221 VID--TEEKRIKALKWT <u>K</u> IEDVWGIGR <u>Q</u> ISKRLQAIKVFTAYDFVNLDNYVKTNFSVVE 220 DLT--HIDRQRKLLALVA <u>E</u> DVWGIGRRISKKLNLMGIETALDAESSLWVIRKHFNVL 221 . : : . * : * : . : * : : : . : : : . : : : . : : : . : : : . : : : . : : :

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.

Fig. 2F Viral polymerases

sp P20311 DPOL_BPT3	RAAFGAEH-HLDGITGKPWVQAGIDASGLELRLCAHFMARFDN-----GEYAHEILNG	503
sp P04415 DPOL_BPT4	KKMFAEEM-NAAEAIKKIIMKGAGSCSTKPEVERYVKFSDDFLNE----LSNYTESVLNS	537
sp P19822 DPOL_BPT5	VAPPGYRVIADLTAEVYYAVALSGD-RNMQQ-VFINMRNEPDKYPDFHSNIAHMVFKL	547
sp P00581 DPOL_BPT7	RAAFGAEH-HLDGITGKPWVQAGIDASGLELRLCAHFMARFDN-----GEYAHEILNG	503
	* . : . * . : : . : . : .	
sp P20311 DPOL_BPT3	DIHTKNQMAAELPTRDNA <u>KTFIYGFYGA</u> G-----	533
sp P04415 DPOL_BPT4	LIEECEKAATLANTNQLNR <u>KLILINSLYGA</u> LGNIHFRYYDLRNATAITIFGQVGIQWIARK	597
sp P19822 DPOL_BPT5	QCERPDVKKLFPALRQAA <u>KAITFGILYGS</u> G-----	577
sp P00581 DPOL_BPT7	DIHTKNQIAAELPTRDNA <u>KTFIYGFYGA</u> G-----	533
	. : . : : * * :	
sp P20311 DPOL_BPT3	-DEKIGQIVG-----AGKERGKELK-----	552
sp P04415 DPOL_BPT4	INEYLNKVC G TNDED F IAAGDTDSVVCVDK V EKG L DRFKEQNDLVEFMNQFGKKME	657
sp P19822 DPOL_BPT5	-PAKVAHSVNE-----ALLEQAAKTG-----	597
sp P00581 DPOL_BPT7	-DEKIGQIVG-----AGKERGKELK-----	552
	: : . . : : :	
sp P20311 DPOL_BPT3	-----KKF	555
sp P04415 DPOL_BPT4	PMIDVAYRELCDYMNNREHLMHMDREAI <u>SCPPLGSKGVGFWKAKRYALNVYDMEDKRF</u>	717
sp P19822 DPOL_BPT5	-----EPF	600
sp P00581 DPOL_BPT7	-----KKF	555
	: *	
sp P20311 DPOL_BPT3	LE-----NTPA--IAALRESIQQT--LVESSQWVAGEQQ-----	585
sp P04415 DPOL_BPT4	AEPHLKIMGMETQQSSTPKAVQEALEE <u>ESIRRI</u> --LQE G EESVQ-EYYKNFEKEYRQLDY	773
sp P19822 DPOL_BPT5	VE-----CTV--ADAKEYIETYFGQFPQLKRWIDKCHD-----	631
sp P00581 DPOL_BPT7	LE-----NTPA--IAALRESIQQT--LVESSQWVAGEQQ-----	585
	* * . * . : : . : :	

Fig. 2F. CLUSTAL format for T-COFFEE Version of polymerases of viral origin

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

Fig. 2G Alpha polymerases (Eukaryotes)

sp P33609 DPOLA_MOUSE	-----NLPLVKGADDEQVFQFYWL DAYEDPYN --QPGVVFLFGKVWIESV-----	374
sp P28040 DPOA_SCHPO	PISKVSPSDVT--EEDGSLFFFWM DYTE --MYGS I LC L FGK V YDKATK-----	356
sp P26019 DPOLA_DROME	VLT-TEQDSTI--SSDQQQLRFWYWEAYEDPVK--MPGEVFLFGRTAD-----	388
sp P13382 DPOA_YEAST	-VNVKKEDVVD--PETDTFQMF WLDYCE --VNNTL I LF G KV K LKD-D-----	377
sp P09884 DPOLA_HUMAN	-----HLPLVKGADEEQVFQFYWL DAYEDQYN --QPGVVFLFGKVWIESA-----	370
sp O89042 DPOLA_RAT	-----HLPLVKGADDEQVFQFYWL DAYEDPYN --QPGVVFLFGKVWIESA-----	377
sp Q9DE46 DPOLA_XENLA	-----HLPLVTGADGSQVFRFYWL DAYEDQYS --QPGVVFLFGKVWIESA-----	370
sp Q9FHA3 DPOLA_ARATH	--SEGKTEFDL--DADGSLRFFI L DAYEEAFG--ASMG T I L FGK V KM-----	372
sp Q54SV8 DPOLA_DICDI	IMPKNSDL IKM --NM D GSLDF FL LTTEED-----KQGR I ILFGK V KLQASKSNKPGGGG	415
sp O00874 DPOLA_LEIDO	--A--VPYKS--GEMTEGLFYWF DAREQPH TL S V D PGS L LF G KMAVEKNG-----	274
sp P27727 DPOLA_TRYBB	--HASTDFLN--GC P KALFYWFDAKEQPH TL -TAPG T ILL F GK V CMNEE-----	272
sp Q94636 DPOLA_OXYNO	MSNQQDYPLPL--NEDGTLSFYWIDAHEENN---GADLF V FGK I YQHEE-----	358
sp O48653 DPOLA_ORYSJ	TGVDESSEFEL--KDGALPFYILDAYEE PFG --ANSGTVYLF G KVEV-----	381
	: : * : : : : ** :	

<p>sp P33609 DPOLA_MOUSE sp P28040 DPOA_SCHPO sp P26019 DPOLA_DROME sp P13382 DPOA YEAST sp P09884 DPOLA_HUMAN sp O89042 DPOLA_RAT sp Q9DE46 DPOLA_XENLA sp Q9FHA3 DPOLA_ARATH sp Q54SV8 DPOLA_DICDI sp O00874 DPOLA_LEIDO sp P27727 DPOLA_TRYBB sp Q94636 DPOLA_OXYNO sp O48653 DPOLA_ORYSJ</p>	<p>S-EYLEVRYSAEV----PQLPQNLKGETFSHVFGNTSSLEIFLMNRKIKGPCWLEVKN 507 A-DYLEVIYSYSP----ALPTDLTGSSFSHVFGNTALFEGFVLSSRRVMGCPWLKIQQ 484 C-DYLEVHYDGKKPL---PNLSADKKYNSIAHIFGATTNALEPFLLDRKIKGPCWLQVTG 523 S-DYLVKLLPYQTPKSSRDTIPSDLSSDTFYHVFGGNSNIFESFVFIQNRMGCPWLDIKG 501 S-EYLEVKYSAEM----PQLPQDLKGETFSHVFGNTSSLEIFLMNRKIKGPCWLEVKS 503 S-EYLEVRYSAEV----PQLPQNLKGETFSHVFGNTSSLEIFLMNRKIKGPCWLEVKN 510 S-EYLEVRYSAADS----PQLPQDLKGETFSHVFGNTSSLEIFLMNRKIKGPCWLEIKS 503 EQYVLKINYSFKD----RPLPEDLKGESFSALLGSHTSALEHFILKRKIMGPCWLKISS 522 QHYVWKLSYPSNQM----VFPNDIKGSTFRCAYGITSSPVEIFLIKRKIMGPTWLTVSG 569 KTQWAKLRYPGRY----PPLNAKGPFRHILIIMGAASSSLLEIFLIJKRKLKGPSFLRISG 395 KNRWAKLVLVPGRY----PPFPNKGGLTHVQVVVGASRSLEIFLIJKRKLKGPSYLEIEH 393 H-QFVKIKYDATF----PSLPSTVQGNTFECIFGANQSMLESFILKRKIRGCPWMТИRN 491 EQYVLKINYPYKD----PALPTDLRGQHFHALLGTNNSALEILLIKRKRKIKGPSWLSISK 530</p>
	:: : . * . * :: : . : ** :: :
<p>sp P33609 DPOLA_MOUSE sp P28040 DPOA_SCHPO sp P26019 DPOLA_DROME sp P13382 DPOA YEAST sp P09884 DPOLA_HUMAN sp O89042 DPOLA_RAT sp Q9DE46 DPOLA_XENLA sp Q9FHA3 DPOLA_ARATH sp Q54SV8 DPOLA_DICDI sp O00874 DPOLA_LEIDO sp P27727 DPOLA_TRYBB sp Q94636 DPOLA_OXYNO sp O48653 DPOLA_ORYSJ</p>	<p>VLPLALQITNIAGNIMSRTLMGGRSERNEFLLHAFYENNYIVPDKQIF-RKPQQKLG-- 821 MLQLSKNLTNIAGNNSWARTLTGTRAERNEYILLHEFKKNGYIVPDKQQSIRRRAEAFG-- 808 IMPLALQITNICGNTMTRLQGGRSERNEFILLHAFYENNYIVPDKQPV-SKRGAG-- 838 LLTLLTKQLTNLAGNAWAQTGLGGTRAGRNEYILLHESRKNYIVPDKEGN-RSRAQKQR-- 826 VLPLALQITNIAGNIMSRTLMGGRSERNEFILLHAFYENNYIVPDKQIF-RKPQQKLG-- 817 VLPLALQITNIAGNIMSRTLMGGRSERNEFILLHAFYENNYIVPDKQIF-RKPQQKPG-- 824 VLPLALQITNIAGNVMMSRTLMGGRSERNEYIILLHAFTEENNIVPDKPVF-KKMQQTTV-- 817 VPLPLTQLTNISGNLWGKTLQGARAQRIEYYLLHTFHSSKKFILPDKISQ-RMKEIKSSKR 861 VFPLTKQLTNLAGNQWDKSLKSNRAERIEYILLHNFEKKYLLIPDKIYQ-KSSSSG-G-- 876 VIRLTKRLTTIAGNWLWSRTLFGARSERIEYILLHTFEDLKFITPDRYVQ-NFKGRD-- 719 VIPLTKRLTLLAGNWLWSRTLYGARSERIEYILLHAFHNLKFVTPDKKKR-DLKRGRE-- 717 VIPLTKQLTNISGNLWFRLSLQNARAERNEILLHEFKKKKVLPDKQLN-AKDLKKNM- 818 VLPLTRQLTNISGNLWGKTLQGSRAQRVEYILLHAFHARKTFVPDKFAR-S-KEFNSTKR 869</p>
	:: * : . * : ** : * : * : *** * : : ** :
<p>sp P33609 DPOLA_MOUSE sp P28040 DPOA_SCHPO sp P26019 DPOLA_DROME sp P13382 DPOA YEAST sp P09884 DPOLA_HUMAN sp O89042 DPOLA_RAT sp Q9DE46 DPOLA_XENLA sp Q9FHA3 DPOLA_ARATH sp Q54SV8 DPOLA_DICDI sp O00874 DPOLA_LEIDO sp P27727 DPOLA_TRYBB sp Q94636 DPOLA_OXYNO sp O48653 DPOLA_ORYSJ</p>	<p>-----DEDEEIDGDTNKKGRKKATYAGGLVLDPKVGFYDKFIILLLDFNS 867 -----AED---G-LQEESLGKKDKYKGGLVFEPOKGLYETCILVMDFNS 849 -----DTDATLSGA-DATMQTKKAAYAAGGLVLEPMRGLYEKYVVLLMDFNS 883 -----QNE---E-NADAPVNSKKAKYQGGLVFEPKGLHKNYVVLMDFNS 867 -----DEDEEIDGDTNKKGRKKAAAGGLVLDPKVGFYDKFIILLLDFNS 863 -----DEDEEIDGDTNKKGRKKAAAGGLVLDPKVGFYDKFIILLLDFNS 870 -----EDNDMDGTDQN-KNKSRRKKAAYAAGGLVLEPKVGFYDKFIILLLDFNS 862 RMDYAPEDRNVDELDADILTLENDPSKGSKTKKGPAAGGLVLEPKRGLYDKYVVLLLDFNS 921 -----G-----GGAKDKDNHAAYSGGLVLDPKIDFYDRYVVLLLDFNS 913 -----D-----EEEEDGKRKAKYQGGMVLDPKCGLYSDYIVLLLDFNS 756 -----D-----DDEGKRKTKYQGGMVLEPKSGLYSEYIVLLLDFNS 753 -----FADEYEEDGDKTKGGKRKAAYAAGGLVIEPKAGFYDNIVLLLDFNS 864 KMNPDTEARPDEADPSIDDEGHHDQGKTKGSPSYAGGLVLEPKKGLYDKYVVLLLDFNS 929</p>
	* : *** : * : * : * : . : : : : **** :
<p>sp P33609 DPOLA_MOUSE sp P28040 DPOA_SCHPO sp P26019 DPOLA_DROME sp P13382 DPOA YEAST sp P09884 DPOLA_HUMAN sp O89042 DPOLA_RAT sp Q9DE46 DPOLA_XENLA sp Q9FHA3 DPOLA_ARATH sp Q54SV8 DPOLA_DICDI sp O00874 DPOLA_LEIDO sp P27727 DPOLA_TRYBB sp Q94636 DPOLA_OXYNO sp O48653 DPOLA_ORYSJ</p>	<p>LYPSIIQEFNICFTTVQRTVSEVQKAT-----E-DEE--QEIQIP----- 903 LYPSIIQEQYNICFTTVDRSPSNS-----DS-DDQIP----- 879 LYPSIIQEQYNICFTTVQQPVDA-----DELP----- 909 LYPSIIQEFNICFTTVDRNKE-----D-I DELP----- 894 LYPSIIQEFNICFTTVQRVASEAQKVT-----E-DGE--QEIQIP----- 899 LYPSIIQEFNICFTTVQRVASETLKAT-----E-DEE--QEIQIP----- 906 LYPSIIQEQYNICFTTVHREAPSTQK-----GED--QDEIP----- 895 LYPSIIQEQYNICFTTIPRSED-G-----VP----- 945 LYPSIIQEQYNVCFTTINRVKRDD-----GK--WEE-A----- 942 LYPSLIQEFNICFTTVDRESGSE-----IDVPPPENLICASCAA 795 LYPSLIQEFNVCYTTDRDENTV-----S-----AEVPPPESLICSLCRA 793 LYPSIIQEQYNLCFTTVNRRTKNFDGSEMKNQYKKGENGEVEEDIEEA----- 912 LYPSIIQEQYNICFTTVDRSADGN-----VP----- 954 **** : * : * : * : * : :</p>
	**** : * : * : * : * : :

sp P33609 DPOLA_MOUSE	-ELPDPN--LEMGILPREIRKLVERRKQVKQLMKQQQDLNPDLVLYDQYDIRQKA	ALKLTANSM	960
sp P28040 DPOA_SCHPO	-DTPSAS--ANQGIFPRILIANLVERRRQIKGLLKDNSATPTQRLQWDIQQQAA	ALKLTANSM	936
sp P26019 DPOLA_DROME	-TLPDSK--TEPGILPLQLKRLVESRKEVKLMAAPDLSPELQMRYHIRQMA	ALKLTANSM	966
sp P13382 DPOA_YEAST	-SVPPSE--VDQGVLPRIILANLVDRRRREVKKVMKTET--DPHKRVQCDIRQQAA	ALKLTANSM	950
sp P09884 DPOLA_HUMAN	-ELPDPS--LEMGILPREIRKLVERRKQVKQLMKQQQDLNPDLVLYDQYDIRQKA	ALKLTANSM	956
sp O89042 DPOLA_RAT	-ELPDPS--LEMGILPREIRKLVERRKQVKQLMKQQQDLNPDLVLYDQYDIRQKA	ALKLTANSM	963
sp Q9DE46 DPOLA_XENLA	-ELPHSD--LEMGILPREIRKLVERRRHHVKQLMKQPDLNPDLVLYDQYDIRQKA	ALKLTANSM	952
sp Q9FHA3 DPOLA_ARATH	-RLPSS---QTPGILPKLMEHLVSIKRSVKKMKTET--GLKYWELDIRQQAA	ALKLTANSM	999
sp Q54SV8 DPOLA_DICDI	-MPPPSS--IEKGILPKVLHGLVSKREIKKRMEQEKK-NKI IKAQYDIQQQAV	ALKLTANSM	998
sp O00874 DPOLA_LEIDO	AGLSAPC--LHKCVPKVIKSLVDSRREVKRLMKIEK-DANNLALLEIRQKA	ALKLTANSM	852
sp P27727 DPOLA_TRYBB	AGLPSPC--LHKCILPKVIRGLVDSREITKRMMSKEK-DPGNLAMLEIROLA	ALKLTANSM	850
sp Q94636 DPOLA_OXENO	-DLPDFNVLKDADLPMVLRDLVQKRKAVKDKMKTEK-DHVKLSQLERQKA	ALKLTANSM	970
sp O48653 DPOLA_ORYSJ	-NLPAS---KTTGVLPPELLKSLVERRRMVKSWLKTA--GLKRQQFDDIQQQAA	ALKLTANSM	1008

. : * : **. * : : * : . : * : * : *** ****

sp P33609 DPOLA_MOUSE	YGCLGFYSYSRFYAKPLAALVTYKGREIILMHTKDMVQKM--NLEVIYGDTSIMINTNST-	1017
sp P28040 DPOA_SCHPO	YGCLGYTKSRFYARPLAVLITYKGREALMNTKELADQM--GLOVIYGDTSVMIINTNVT-	993
sp P26019 DPOLA_DROME	YGCLGFAHSHRFQAQHIAALVTHKGREIILTNTQLVQKM--NYDVVYGDTSVMIINTNIT-	1023
sp P13382 DPOA_YEAST	YGCLGYVNSRFYAKPLAMLVTNKGREIILMNTRQLAESM--NLLVVVYGDTSVMIIDTGCD-	1007
sp P09884 DPOLA_HUMAN	YGCLGFYSYSRFYAKPLAALVTYKGREIILMHTKEMVQKM--NLEVIYGDTSIMINTNST-	1013
sp O89042 DPOLA_RAT	YGCLGFYSYSRFYAKPLAALVTYKGREIILMHTKEMVQKM--NLEVIYGDTSIMINTNST-	1020
sp Q9DE46 DPOLA_XENLA	YGCLGFYSYSRFYAKPLAALVTHQGREIILHTKEMVQKM--NLEVIYGDTSIMINTNCN-	1009
sp Q9FHA3 DPOLA_ARATH	YGCLGFNSNSRFYAKPLAELITLQGRDLQRTVDLVQN-HLNLEVIIYGDTSIMIHSGLD-	1057
sp Q54SV8 DPOLA_DICDI	YGCLGFSSHRSRFYALPLAELVTRKGRENLQKGASIVNK--CYDVIYGDTSLSMIYTGVG-	1055
sp O00874 DPOLA_LEIDO	YGCLGFSEYSRFYAQPLAELVTRQGRRALQSTVLDLIPQLNPNSLEVIIYGDTSVMIQTGIKN	912
sp P27727 DPOLA_TRYBB	YGCLGFSEYSRFYAQPLAELVTRQGRRALQNTVELIPQISPSIRVIVIYGDTSVMIQTGIKD	910
sp Q94636 DPOLA_OXENO	YGCLGFSSSRFHQAQIAALITRTGRETLLRTKDAIAEN-KLGFNVVYGDTSIMINTGSN-	1028
sp O48653 DPOLA_ORYSJ	YGCLGFNSNSRFYAKPLAELITLQGREIILQNTVDLVQN-NLNLEVIIYGDTSIMIHTGLD-	1066

*****: :*** * : * : * : ** * : .. : * : *****: : ..

sp P33609 DPOLA_MOUSE	NLEEVFKLGNGVKSEVNKLYK--LLEIDIDAVFKSLLLLKKKKYAA	LVVEPT-SDGNYIT	1074	
sp P28040 DPOA_SCHPO	DKNHALRIGNEFKEKVNERYS--KLEIDIDNVYQRMLLHAKKKYAAQLDSQ---G--KP	1046		
sp P26019 DPOLA_DROME	DYDVKYKIGHNIQCSVNKKLYK--QLELDIDGVFGCCLLKKKKYAAIKLSKD-SKGNLRR	1080		
sp P13382 DPOA_YEAST	NYADAIIKIGLGFKRLVNERYR--LLEIDIDNVFKLLLHAKKKYAA	LTVNL-KNGNGIT	1064	
sp P09884 DPOLA_HUMAN	NLEEVFKLGNGVKSEVNKLYK--LLEIDIDGVFKSLLLLKKKKYAA	LVVEPT-SDGNYVT	1070	
sp O89042 DPOLA_RAT	NLEEVFKLGNGVKSEVNKLYK--LLEIDIDGVFKSLLLLKKKKYAA	LVVEPT-SDGNYIT	1077	
sp Q9DE46 DPOLA_XENLA	NLEEVFKLGNGRKSEINKSYK--LLEIDIDGIFKSSLKKKKYAA	LTVFPT-GDGKYVT	1066	
sp Q9FHA3 DPOLA_ARATH	DIEEVKAISKVVIQEVNNKKYR--CLKIDCDGIYKRMLLLKKKKYAAVKLQ-F-KDGKPCE	1113		
sp Q54SV8 DPOLA_DICDI	TFNEAETIGKEIQKKINDQYRG SVM EIGLDGIFKRLLLFKKKYAA	CLKEFRI-DSTTTKC	1114	
sp O00874 DPOLA_LEIDO	DIKAVRDLGLDLKAKINKRYQ--SLEIDIDGVFRAI	LLLKKKKYAA	LTVDW-QGEGKTY	969
sp P27727 DPOLA_TRYBB	DIVKVRNLGFEIKGKVNRQRYQ--SLELDIDGVFRAMLLLRKKKYAA	LSVVDW-QGEGKVV	967	
sp Q94636 DPOLA_OXENO	QLQQALEM GKR LKG EVN CLYK--CLEIEIDGVFKSLLLKKKKYAA	ALKYENFL SPAEV KV	1086	
sp O48653 DPOLA_ORYSJ	DISRAKGIAAGKVIQEVNNKKYR--CLEIDL DGIYKRMLLLKKKKYAA	IKVA-L--DGSLRE	1121	

. : . : * * : :: : * : : * : : *** *****:

sp P33609 DPOLA_MOUSE	KQELKGLD IVRRDWCDLAKDTGNFVIGQILSDQSRDTIVENIQKRLIEIGENVLNGSV-P	1133
sp P28040 DPOA_SCHPO	NLDVKGLDMKRREFCTI LAKEASKFCLDQILS GELTETVIENIH SYLMDFSEKMRNGKF-P	1105
sp P26019 DPOLA_DROME	EQEHKGLD IVRRDWSQIAVMVGKAVLDEVLSEKPLEEKILD DAVHAQLEKIKTQIAEGVV-P	1139
sp P13382 DPOA_YEAST	VLEVKG GLDMKRREFCPLSRDVSIHV LNTL ISDKDPEEALQEVYD LED IRIKVETNNI-R	1123
sp P09884 DPOLA_HUMAN	KQELKGLD IVRRDWCDLAKDTGNFVIGQILSDQSRDTIVENIQKRLIEIGENVLNGSV-P	1129
sp O89042 DPOLA_RAT	KQELKGLD IVRRDWCDLAKDTGNFVIGQILSDQSRDTIVENIQKRLIEIGENVLNGSV-P	1136
sp Q9DR46 DPOLA_XENLA	KQELKGLD IVRRDWCELAQAGNYVISQ SQLSDQPRDS IVENIQK KLTEIGENVNGTV-P	1125
sp Q9FH43 DPOLA_ARATH	DIERKG VIVRRDW WSLLSK EIGDLCLSKI LYGGSCEDVVEA HNELMKIKEEMRNGQV-A	1172
sp Q54SV8 DPOLA_DICDI	ERENKG IDIVRRD YCDLT KDIQWV LNLL IGGEEKIALFSLIK EY LE SVQQ QIKDNTL-A	1173
sp O00874 DPOLA_LEIDO	KKEVK GGLDMV RRDW CPLSKVC CD SLSRVLNAE GSED G SE DILDFV IKY MKGVAQDVRSGNVYP	1028
sp P27727 DPOLA_TRYBB	KREVKG GLDMV RRDW CPLS QH VSDAVLKR LN NAEGGED D ILDFV IKY MKGVAQDVRSGNVYP	1027
sp Q94636 DPOLA_OXENO	VQEMKG GLDMV RRDW CPLSKRVGRY V LDQILSGK Q REEVV LN NLF SN IGNELKEGTI-K	1145
sp O48653 DPOLA_ORYSJ	NIERKG GLDMV RRDW WSLLSK E IGDFCLNQI LSGG SCDDVIESIH SSLV QVQEQMRRGGQT-E	1180
	: * : * : : * : . * : : * . : : . : . : .	

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)

gi 4505931 ref NP_002681.1	-----AYRKAAS VIAK -----	48
sp P06766 DPOLB_RAT	-----AYRKAAS VIAK -----	48
tr Q66TP6 Q66TP6_XIPMA	-----AYRKAAS VIAK -----	48
sp Q27958 DPOLB_BOVIN	-----AYRKAAS VIAK -----	48
sp P06746 DPOLB_HUMAN	-----AYRKAAS VIAK -----	48
tr D0EW68 D0EW68_PONPY	-----AYRKAAS VIAK -----	27
sp Q5UQR0 DPOL_MIMIV	KH KFYNFRNKQLYNFLLVFKSHTAMKEFSSILAR PEAKGLTNKPMILYQRYESNIEPHI	167
	* : : : * : * : * :	
gi 4505931 ref NP_002681.1	-----NHHQRI G LYF GDFE K-----	148
sp P06766 DPOLB_RAT	-----NHHQRI G LYF GDFE K-----	148
tr Q66TP6 Q66TP6_XIPMA	-----NHHQ KIG LYF GDFE K-----	148
sp Q27958 DPOLB_BOVIN	-----NHHQRI G LYF GDFE K-----	148
sp P06746 DPOLB_HUMAN	-----NHHQRI G LYF GDFE K-----	148
tr D0EW68 D0EW68_PONPY	-----NHHQRI G LYF GDFE K-----	127
sp Q5UQR0 DPOL_MIMIV	FAKDDMH HTKIN-EYFEEGDPKKIRQIAKYCL DKCLVNLLAKLEII NSVGMAKVCHV	645
	* : * : * : * : * : * :	
gi 4505931 ref NP_002681.1	---RIPRE---EMLQM D IVLNEV KKVDSEYIAT ---	176
sp P06766 DPOLB_RAT	---RIPRE---EMLQM D IVLNEV KKLDPEYIAT ---	176
tr Q66TP6 Q66TP6_XIPMA	---RIPRV---EMEKME V L LGEI KKIDPEYI GT ---	176
sp Q27958 DPOLB_BOVIN	---RIPRE---EMLQM D IVL SEVKKVDSEYIAT ---	176
sp P06746 DPOLB_HUMAN	---RIPRE---EMLQM D IVLNEV KKVDSEYIAT ---	176
tr D0EW68 D0EW68_PONPY	---RIPRE---EMLQM D IVLNEV KKVDSEYIAT ---	155
sp Q5UQR0 DPOL_MIMIV	LSKNL NEEKSKQINKMELNTKNLISKVFSKYLITEQ QREELIV LEKRAKRSVNAEKAK	945
	: . : : * : . * : * : * : * : * : .	
gi 4505931 ref NP_002681.1	-----VCGSF R -----GAESS G DMDV L I THPSFTSESTK -OPK L I HQVVE 216	
sp P06766 DPOLB_RAT	-----VCGSF R -----GAESS G DMDV L I THPSFTSESSK -OPK L I HRVVE 216	
tr Q66TP6 Q66TP6_XIPMA	-----VCGSY R -----GAASS G D IDIDILL THPNFTS QTEK -OPK L I HAVVD 216	
sp Q27958 DPOLB_BOVIN	-----VCGSF R -----GAESS G DMDV L I THPSFTSES A K -OPK L I HRVVE 216	
sp P06746 DPOLB_HUMAN	-----VCGSF R -----GAESS G DMDV L I THPSFTSE STK-OPK L I HQVVE 216	
tr D0EW68 D0EW68_PONPY	-----VCGSF R -----GAESS G DMDV L I THPSFTSE STK-OPK L I HQVVE 195	
sp Q5UQR0 DPOL_MIMIV	EAYVWGFFMGDG CGSYQITKNGIKY SWALNN QLDV LN KCKKY LEETENI QFKI LD TMKS	1245
	* * : : * .. * : * : * : . . : : * : * : .	

gi 4505931 ref NP_002681.1	TKFMGVCQLPSKND--EKEY <u>PHRRRIDIR</u>	----- 258
sp P06766 DPOLB_RAT	TKFMGVCQLPSEN--ENEY <u>PHRRRIDIR</u>	----- 258
tr Q66TP6 Q66TP6_XIPMA	TKFMGVCQLQDTDDDEEEH <u>LHRRRIDIR</u>	----- 260
sp Q27958 DPOLB_BOVIN	TKFMGVCQLPSKND--EKEY <u>PHRRRIDIR</u>	----- 258
sp P06746 DPOLB_HUMAN	TKFMGVCQLPSKND--EKEY <u>PHRRRIDIR</u>	----- 258
tr D0EW68 D0EW68_PONPY	TKFMGVCQLPSKND--EKEC <u>PHRRRIDIR</u>	----- 237
sp Q5UQR0 DPOL_MIMIV	TENMGC----- <u>RRCDIKGKISAQCLFYLLKSLGYNVSI</u> NIRS DKNQIYR	1349
*	***	
gi 4505931 ref NP_002681.1	CGVLYFT-	----- 273
sp P06766 DPOLB_RAT	CGVLYFT-	----- 273
tr Q66TP6 Q66TP6_XIPMA	CGVLYFT-	----- 275
sp Q27958 DPOLB_BOVIN	CGVLYFT-	----- 273
sp P06746 DPOLB_HUMAN	CGVLYFT-	----- 273
tr D0EW68 D0EW68_PONPY	CGVLYFT-	----- 252
sp Q5UQR0 DPOL_MIMIV	VGLLFEKSPDKYFLKSMGIVLKRRDNAPIVKIVVGGIIDNILKNRDIDKAIEYTKIVLDK	1528
*	**	

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	-----IRAEDLNKNTFVVKVPHPDGPSARVTNCMTKSCLGFF	567
sp Q12704 DPOG_SCHPO	-----DTKLDYNNYIFVKVPHKDGEARCGSPLSKSYQRYF	574
sp P15801 DPOG_YEAST	-----NLGLQC-TGVLFKVPHPNPPTFNCTNLLTAKSYNHFF	576
sp Q9Y767 DPOG_NEUCR	-----RLRMDV-DHKYFKLPHKDGPARNACVNPMAGKYL PYF	626
sp Q92076 DPOG1_CHICK	EDGLPE-LVEESSQPSFHNGNPY-NDVNIPGCWFKLPHKDGNENN VGS PFAKDFLP RM	186
sp P54098 DPOG1_HUMAN	PLALTARGPKDTQPSYHHNGNPY-NDVDI PG CWF KLP HKDGNCSNV GSPFAKDFLP KM	774
sp Q27607 DPOG1_DROME	YKKL---SQKQQRLETQYQGSGVWCNKVLLDCCFFKLKLP HRN GP SFRVG NPLSKDFLN KF	693
sp P54099 DPOG1_MOUSE	PLVLPAACPKSSQPTYHHNGNPY-NDVNIPGCWFKLPHKDGNNN NVGSPFAKDFLP KM	753
sp Q91684 DPOG1_XENLA	PVKL---EMEFDSLDPDNHHGNSPC-GDVNVSGCW FYKLPHKDGNANN NVGSPFAKDFLP KM	744
*	**** : * . . : : * . : * . : * . : *	
sp Q01941 DPOG_PICPA	EKGFLNSQYP---LAKDALQMAVASSYWTSSRERIMQNQFVV F-----	607
sp Q12704 DPOG_SCHPO	EEGILQSDYE---VAKKALEMMSASC SYWSSARDRIRQS QMVW DKDAELG-V-----P	622
sp P15801 DPOG_YEAST	EKGVLKSESE---LAHQALQINSSG SYWMSARERIQS QSFV FPSC KFPNEFQSLS A KSSLN	633
sp Q9Y767 DPOG_NEUCR	EKGILSSEY---YAKEALEMNASC SYWISARERIKNQM VYEDQLPPS-QRFVNKD-AD	681
sp Q92076 DPOG1_CHICK	EDGTLRAAVG-RTHGTRALEINKMVS FWRNAHKRVSSQV VWLKGEL P-RAVTRHP-A Y	243
sp P54098 DPOG1_HUMAN	EDGTLQAGPG-GASGPRALEINKMIS FWRNAHKR I SSQMV VWL PRSALP-RAVIRHP-DY	831
sp Q27607 DPOG1_DROME	AENVLSSGDPS CQAA RVIDA RMM SY WRRN NDRIM QGM QMV VWL DSQQLP-NEFTGEK---	749
sp P54099 DPOG1_MOUSE	EDGTLQAGPG-GASGPRALEINKMIS FWRNAHKR I SSQMV VWL PRSALP-RVVTRHP-A F	810
sp Q91684 DPOG1_XENLA	EDGTLQASTG-DSSATRALEINKMIS FWRNAHKR I SSQMV VWMK NELH-RTITRDP-E F	801
.. * : . . . : : * : . : * : . : * . : *		
sp Q01941 DPOG_PICPA	----DDMGYI LPQII PMGTITRRAV E-----	663
sp Q12704 DPOG_SCHPO	S-SVDGFGI ILPCIIPM GTVTRRAV E-----	681
sp P15801 DPOG_YEAST	NEKTN DLAI IIPKIV PMGTITRRAV E-----	693
sp Q9Y767 DPOG_NEUCR	SN-TPIGGFVLPQVIPMGTITRRAV E-----	740
sp Q92076 DPOG1_CHICK	SE-EEDYGAILPQV VTA-----	302
sp P54098 DPOG1_HUMAN	DE-EGLYGAILPQV VTA-----	890
sp Q27607 DPOG1_DROME	CQ-PIAYGAICPQV VAC-----	808
sp P54099 DPOG1_MOUSE	DE-EGHYGAILPQV VTA-----	869
sp Q91684 DPOG1_XENLA	DE-ENKYGAILAQV VSA-----	860
. . . : : * * * * : : * * * * : : * * * * : : * * * . . . :		

sp Q01941 DPOG_PICPA	VDSEELWIASLIGDSVF-KIHGGTAIGWMTLEGTKNE	GTDLHSKTAKILGISRNE	AKIFN	722
sp Q12704 DPOG_SCHPO	VDSEELWIVALMGDSQF-RLHGATALGMMTLEGKSE	GTDLHSKTAAILGVSRDS	AKVFN	740
sp P15801 DPOG_YEAST	VDSEELWIASLVGDSIF-NVHGGTAIGWMCLEGTKNEG	GTDLHTKTAQILGCSRNE	AKIFN	752
sp Q9Y767 DPOG_NEUCR	VDSEELWIASVVGDAFT-KLHGGNAIGFMTLEGTSQG	GTDLHSRTASILGITRND	AKVFN	799
sp Q92076 DPOG1_CHICK	VDSQELWIAAVLGEAHFAGMHGCTAFGWMTLQGKKS	GTDLHSKTAATVGISRE	HAKVFN	362
sp P54098 DPOG1_HUMAN	VDSQELWIAAVLGEAHFAGMHGCTAFGWMTLQGRKS	GTDLHSKTAATVGISRE	HAKIFN	950
sp Q27607 DPOG1_DROME	VDSQELWIAAVLGEAHFAGIPLGWMTLSGSKSNGSDM	HHSITAKAVGISRD	HAKVIN	868
sp P54099 DPOG1_MOUSE	VDSQELWIAAVLGEAHFAGMHGCTAFGWMTLQGRKS	GTDLHSKTAATVGISRE	HAKIFN	929
sp Q91684 DPOG1_XENLA	VDSQELWIAAVLGEAHFAGIPLGWMTLSGSKSNGSDM	HHSITAKAVGISRD	HAKVIN	920
	.*..::*:::	** .. :* * .*. * .:::*	** :* ::: **:::*	
sp Q01941 DPOG_PICPA	YGRIYAGAIKFTTLLKKFNPALSDEAKATANALYTAT	TKGKIS-GR-----		767
sp Q12704 DPOG_SCHPO	YGRLYGAGLKHTTLLLMQMNPNTLKTAEAKLKYAST	TKGVSKM-----		786
sp P15801 DPOG_YEAST	YGRUYGAGAKFASQLLKRNFNPSLTDEETKIANLYENT	TKGKT-KR-----		797
sp Q9Y767 DPOG_NEUCR	YGRUYGAGLKFASQLLRQFNPSLTEAETTAIATKLYDA	TKGAKTNR-----		845
sp Q92076 DPOG1_CHICK	YGRUYGAGOPFAERLLMQFNHRLTQQQAREKAQ	QOMYAVTKGIRRFLSEEGLVKELE	L	422
sp P54098 DPOG1_HUMAN	YGRUYGAGOPFAERLLMQFNHRLTQQQAREKAQ	QOMYAATKGLRWYRLSDEGEGLVRELNL		1010
sp Q27607 DPOG1_DROME	YARIYAGOLFAETLRLQRNFNPTFSASEAKAKAMK	MFSITKGKRVYRLREEFHDELED--		925
sp P54099 DPOG1_MOUSE	YGRUYGAGOSFAERLLMQFNHRLTRQEEAKA	QOMYAVTKGLRYYRLSADGEGLVQQLNL		989
sp Q91684 DPOG1_XENLA	YGRUYGAGOPFAERLLMQFNHRLTQQQAREKAQ	QOMYAVTKGIRRYYILSKEGEGLVVEELGI		980
	*.**.***. . : * * :* : . : : * : ***			
sp Q01941 DPOG_PICPA	-----Y-----DKKSIWYGGSESIIIFNRLE	AIAEMAH		794
sp Q12704 DPOG_SCHPO	-----SK-RLQEMGLPKLTFWSQGTESFVFNKLE	AMAQLPS		821
sp P15801 DPOG_YEAST	-----SK-----LFKKFWYGGSESILFNKLESIAEQET			825
sp Q9Y767 DPOG_NEUCR	-----KS-----LYKRPFWRGGTESFVNMLEEFAEQER			874
sp Q92076 DPOG1_CHICK	AVDKAEDGTVAQDVQKIQREAMRKSSRRKKWDVV	AHRMWAGGTESSEMFKLLESIALSAS		482
sp P54098 DPOG1_HUMAN	AVDKAEDGTVAQDVQKIQREAMRKSSRRKKWDVV	AHRMWAGGTESSEMFKLLESIALSAS		482
sp Q27607 DPOG1_DROME	PVDRTEGGWISLQDLRKVQRETARKSQW-KKWEVVA	ERAWKGTESEMFNKLESIATSDI		1069
sp P54099 DPOG1_MOUSE	PVDRTEGGWISLQDLRKVQRETARKSQW-KKWEVVA	ERAWKGTESEMFNKLESIATSDI		1069
sp Q91684 DPOG1_XENLA	PVDRTEGGWISLQDLRKVQRETARKSQW-KKWEVVA	ERAWKGTESEMFNKLESIATSDI		1069
	* * :* :* :* :* :* :* :* :			
sp Q01941 DPOG_PICPA	PKTPVLSAGITAPLQKANL--STNNFLTSRINWAI	QSSGVYDYLHLLIISMDYLIKLFID		852
sp Q12704 DPOG_SCHPO	PRTPVLDAGITQALSSKNL--SKNSFMTSRVNWAI	QSSAVDYLHLLLVSMNHLIKKYLE		879
sp P15801 DPOG_YEAST	PKTPVLCGCGITYSLMKKNL--RANSFLPSRINWAI	QSSGVYDYLHLLCCSMEYIICKYNLE		883
sp Q9Y767 DPOG_NEUCR	PRTPVLAGITEALMSRVV--SKGGFLTSRINWAI	QSSGVYDYLHLLIAMDYLTRRFNLA		932
sp Q92076 DPOG1_CHICK	PQTPVLSCHISRALEPVA--AKGEFLTSRNVWV	QSSAVDYLHLLVSMKWLFEYDIN		539
sp P54098 DPOG1_HUMAN	PRTPVLGCCISRALEPSA--VQEEMFTSRNVWV	QSSAVDYLHLLVAMKWLFEEFAID		1126
sp Q27607 DPOG1_DROME	PRTPFLGGGRSRALEADTGPEQRFLPTRINWV	VQSGAVDYLHLLLVSMRWLMLGS--H		1030
sp P54099 DPOG1_MOUSE	PRTPVLCGCISRALEPSV--VQGEFITSRVNVWV	VQSSAVDYLHLLVAMKWLFEEFAID		1105
sp Q91684 DPOG1_XENLA	PKTPVLCRISRALEPTA--VKGEFITSRVNVWV	VQSSAVDYLHLLVAMKWLFEEAYDID		1095
	*.**.**. : . * : * :* :* :* :* :* :			
sp Q01941 DPOG_PICPA	ARLCITVHDEIRYLVKEEDKFRAAYALQISNLWTRAM	FCQQQLGINEVVPQSCAFFSAVDLD		912
sp Q12704 DPOG_SCHPO	ARLSLTIVHDEVRYLSSDKDYRVAFAQLQVANLWTR	AFFCQLGINELPQSVAFFSSVDID		939
sp P15801 DPOG_YEAST	ARLCISIHDEIRFLVSEKDQYRAAMALQISNIWTRAM	FCQQQLGINELPQNCAFFSQVDID		943
sp Q9Y767 DPOG_NEUCR	CRLAITVHDEIRYLAEFPDKYRVAMALQIANLWTR	VMFAQQVGIQDLPQSCAFFSAVDID		992
sp Q92076 DPOG1_CHICK	GRFCISIHDEVRYLQEQDRYRAALALQITNLLTRCM	FAYKLGLNDLPQSVAFFSAVDID		599
sp P54098 DPOG1_HUMAN	GRFCISIHDEVRYLVREREQDRYRAALALQITNLLTR	CMFAYKLGLNDLPQSVAFFSAVDID		1186
sp Q27607 DPOG1_DROME	VRFCLSHDELRYLKEELSPKAALAMHITNLMTR	SRCVSRIGLQLDLPMSV AFFSSVEVD		1090
sp P54099 DPOG1_MOUSE	GRFCISIHDEVRYLVREREQDRYRAALALQITNLLTR	CMFAYKLGLNDLPQSVAFFSAVDID		1165
sp Q91684 DPOG1_XENLA	GRFCISIHDEVRYLVHSKDRYRAALALQITNLLTRCM	FASRLGIQDVPQSVAFFSAVDID		1155
	*.**.**.**. . : * :* :* :* :* :* :* :			

sp Q01941 DPOG_PICPA	FVLRKEVDLDCVTPSNPD-----PIPCGKSLDIYQLLQQEDIKG-A--DFPRT---MHL	960
sp Q12704 DPOG_SCHPO	HVLRKDVMDCVTPSNKV-----PIPPGEELTIESVLEKLEQSG-QS--LEPL---EQI	987
sp P15801 DPOG_YEAST	SVIRKEVNMDCITPSNKT-----AIPHGEALDINQLLDKSNSKL-GKPNLDID---SKV	993
sp Q9Y767 DPOG_NEUCR	HVLRKEVMDCITPSNPI-----PIAHGESIDIFQILEKGDDAKLDDSIVPQSQYAPRL	1046
sp Q92076 DPOG1_CHICK	RCLRKEVTMNCATPSNPTGMEEKYGYIPGRGEALDIYQIIIEITKGSL-E-----	645
sp P54098 DPOG1_HUMAN	RCLRKEVTMDCKTPSNPTGMERRYGIPQGRALDIYQIIIEITKGSL-EK--R-----	1234
sp Q27607 DPOG1_DROME	TVLRLKECTMDCKTPSNPHGLRIGYGYIPQGQSLSVAAEIAKAGGND-VS-----	1137
sp P54099 DPOG1_MOUSE	QCRLRKEVTMDCKTPSNPTGMERRYGIPQGEALDIYQIIIEITKGSL-EK--RKP-----	1215
sp Q91684 DPOG1_XENLA	KCLRKEVTMDCSTPSNPNGMEKRYGYIPQGEALDIYQILKVTKG-----	1199
	: * : : * **** * * : : . . .	

Fig. 2I. CLUSTAL format for T-COFFEE version of mitochondrial γ polymerases (eukaryotes)

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

Fig. 2J Delta polymerases (Eukaryotes)

sp P30316 DPOD_SCHPO	PYFYVKAPVGRPEM---LERFTQDLATCN-----G---GVIDHCIIEMK	176
sp P15436 DPOD_YEAST	NYLYVPAPNSSDANDQEIQINKFVHYLNETF-----D---HAIDSIEVVSK	199
sp P28339 DPOD1_BOVIN	PYFYTPAPPFGFPEH---LSELQRELSAISRDQRG-G---KELTG---PAVLAVELCSR	194
sp P28340 DPOD1_HUMAN	PYFYTPAPPFGFPEH---MGDLQRELNLAISRDSRG-G---RELTG---PAVLAVELCSR	195
sp O54747 DPOD1_RAT	PYFYTPAPPFGGAEH---LSELQRELNAASRDQRG-G---KELSG---PAVLAIELCSR	191
sp P30315 DPOD1_PLAFK	PYFYVEKPDFDNED---LIKLEMNENLNLSQY-----KIYE---KKLIKIEIVKT	187
sp P54358 DPOD1_DROME	PYFYTEAPSQFEEHH---CEKLQKALDQKVIADIRN-N---KDNVQ---EAVLMLVELVK	175
sp P52431 DPOD1_MOUSE	PYFYTPAPPFGFPEH---LSELQQELNAAISRDQRG-G---KELSG---PAVLAIELCSR	193
sp P90829 DPOD1_CAEEL	PHFYFQAPQPGFGEVH---IGTAQSAICNMVAAKRGGSGQAQLPGKVVDNLVHVEIVHG	163
sp Q9LVN7 DPOD1_ARATH	PYFYIACPPGMGPDD---ISNFHQSLERGRMRESNKN-A---KVP---KFVKRIEMVQK	190
sp Q9LRE6 DPOD1_ORYSJ	PYFYVISCPGMGPDD---ISRFHQTLERGRMKDSRN-S---NVP---RFVKRIELVQK	200
sp P46588 DPOD_CANAL	HYFYCPVPKGFE-EN---LTEFTNYLKATF-----DGIERVEITSK	136
sp P97283 DPOD1_MESAU	PYFYTPAPPFGFPEH---LSDLQRELSTAISRDQRG-G---KELSG---PAVLAIELCSR	191
sp O48901 DPOD1_SOYBN	PYFYVICCPGMGPDD---ISHFHQTLEGRMREANRN-S---NVG---KFVRRIEMVQR	157
	: * :	

sp P30316 DPOD_SCHPO	APLRLIMSF DIECA GRKGV-FPDPSIDPVI QIASIVTQYGDST-PFVRNVFCVDTCSQIVG	349
sp P15436 DPOD_YEAST	APLRLIMSF DIECA GRIGV-FPEPEYDPVI QIANVVSIA GAKK-PFIRNVFTLNTCSPITG	370
sp P28339 DPOD1_BOVIN	APLRVLSFDIECA GRKGI-FPEPERDPVI QICSLGLRWGEPE-PFLRLALT LRPCAPILG	364
sp P28340 DPOD1_HUMAN	APLRVLSFDIECA GRKGI-FPEPERDPVI QICSLGLRWGEPE-PFLRLALT LRPCAPILG	365
sp O54747 DPOD1_RAT	APLRVLSFDIECA GRKGI-FPEPERDPVI QICSLGLRWGEPE-PFLRLALT LRPCAPILG	361
sp P30315 DPOD1_PLAFK	PKLRLISFDIECAIKLDGKGFPEAKNDPII QISSILYFQGE PIDNCTKFIFTLLECASIPG	359
sp P54358 DPOD1_DROME	APFRILISFDIECA GRKGI-FPEAKIDPVI QIANMVIRQGERE-PFIRNVFTLINECAPIIG	349
sp P52431 DPOD1_MOUSE	APLRVLSFDIECA GRKGI-FPEPERDPVI QICSLGLRWGEPE-PFLRLALT LRPCAPILG	363
sp P90829 DPOD1_CAEEL	APIRTLSLDIECIGRRGV-FPEAIKDP IIQIANLVKIEGEAE-PFVRNCFVLGTCAPVVG	335
sp Q9LVN7 DPOD1_ARATH	APFRVLSFDIECA GRKGH-FPEAKHD PVI QIANLVTLQGEDH-PFVRNVMTLKSCAPIVG	359
sp Q9LRE6 DPOD1_ORYSJ	APFRILISFDIECA GRKGH-FPEPTHD PVI QIANLVTLQGEQ-PFVRNVMTLKSCSPIVG	369
sp P46588 DPOD_CANAL	APLRVLSFDIECA GRKGV-FPEAEHD PVI QIANVVKQSGESK-PFVRNVFTVNTCSSIIIG	299
sp P97283 DPOD1_MESAU	APLRVLSFDIECA GRKGI-FPEPERDPVI QIANLVTLQGEDQ-PFIRNVMTLKSCSPIVG	361
sp O48901 DPOD1_SOYBN	APFRILISFDIECA GRKGH-FPEPTHD PVI QIANLVTLQGEDQ-PFIRNVMTLKSCSPIVG	326
	. : * : * : * * * * ** : . * * : * * .. : * : : : * : : *	

sp P30316 DPOD_SCHPO	TQVYEFQNQAEMLSSWSKFVRDVDPDVIGYNI	CNFDIPYLDRAKSLRIHNFPLLGRIH	409			
sp P15436 DPOD_YEAST	SMIFSHATEEEMLSWRNFIIKVDPDVIGYNT	TFDIPYLNRAKAKVNDFPFYGRKL	430			
sp P28339 DPOD1_BOVIN	AKVQSYEREEDLLQAQAWSTFIRIMDPDVITGYNI	QNFDLPYIISRAQTLKVPGFPLLGRI	424			
sp P28340 DPOD1_HUMAN	AKVQSYEKEEDLLQAQAWSTFIRIMDPDVITGYNI	QNFDLPYIISRAQTLKQTFPFLGRVA	425			
sp O54747 DPOD1_RAT	AKVQSYEREEDLLQAQAWATFI	LANDDPDVITGYNIQNFDLPYIISRAQTLKVDRFPFLGRVT	421			
sp P30315 DPOD1_PLAFK	SNVIWFNDEKTLLEAWNEFFIIRD	DPFDLTGYNIINFDFLPYILNRGTA	419			
sp P54358 DPOD1_DROME	SQVLCHDKETQMLDKWSAVFREV	DPFDILTGYNNNFDFPYNLRRAHLKVR	409			
sp P52431 DPOD1_MOUSE	AKVQSYEREEDLLQAQAWAD	DPFDLTGYNIQNFDLPYIISRAQALKVDRFPFLGRVT	423			
sp P90829 DPOD1_CAEEL	SNI	IQCVNEKVLEKWAEVREVDPDIITGYN	LNFDLPYILDRAKVLSIPOVSHLGRQK	395		
sp Q9LVN7 DPOD1_ARATH	VDVMSFET	TEREVVLLAWRLD	IIRDLPYIIERAATLGIEEFPFLGRVK	419		
sp Q9LRE6 DPOD1_ORYSJ	TDK	FETEVLLAWRLDFIREVDPDII	IGYNCFKFDLPYIIERAEVLKIVEFPILGRIR	429		
sp P46588 DPOD_CANAL	SQIFEHQ	REEDMLMHKEF	ITKVD	DPDIIIGYNTANFDIPYVLNRAKALGLNDPF	FFGRLK	359
sp P97283 DPOD1_MESAU	AKVQSYEREEDLLQAQWPN	FILAMD	DPFDLTGYNIQNFDLPYIISRAQTLKVDRFPFLGRVT	421		
sp O48901 DPOD1_SOYBN	VDVMPFET	EREVVLLAWRLDFIREVDPDIIIGY	NCFKFDLPYIIERALNLKIAEFPI	GRIR	386	
:	:	:*: * : :	*** : ***	:***:***: *.* . * : . :**		
sp P30316 DPOD_SCHPO	NFFSVAKETTFSSKAYGTRESKTT	SIPGRLQLDMQLQMQRDFKLRSYS	LNACVSQFLGEQ	469		
sp P15436 DPOD_YEAST	TVKQEIKEVFSSKAYGTRET	KNVNIDGRLQLDQQF	IQREYKLRSYTLNAVSAHFLGEQ	490		
sp P28339 DPOD1_BOVIN	GLRSNIRESSFQSRTGRRDSKV	VMSMGRVQMDMLQ	VQMDMLQVLLREYKLRSYTLNAV	SFHFLGEQ	484	
sp P28340 DPOD1_HUMAN	GLCSNIRDSSFQSRTGRRDTK	VVMSMGRVQMDMLQ	VQMDMLQVLLREYKLRSYTLNAV	SFHFLGEQ	485	
sp O54747 DPOD1_RAT	GLRSNIRDSSFQSRTGRRD	SKV	VMSMGRVQMDMLQVLLREYKLRSYTLNAV	SFHFLGEQ	481	
sp P30315 DPOD1_PLAFK	NVASTVKDSSFQS	QVGRGDRDSKV	VMSMGRVQMDMLQVLLREYKLRSYTLNAV	SFHFLGEQ	479	
sp P54358 DPOD1_DROME	NIRSVIKEQMLQSKQMGR	RRENQYVNFEGR	GRVFPFDLLVLLRDYKLRSYTLNAV	SYHFLQEQ	469	
sp P52431 DPOD1_MOUSE	GLRSNIRDSSFQSRTGRR	DSKVISMGRVQMDMLQ	VQMDMLQVLLREHKLRSYTLNAV	SFHFLGEQ	483	
sp P90829 DPOD1_CAEEL	EKGSVVRDAAISSKQMGSR	VNKSIDIHGRI	IFDVLQVVLRDYKLRSYTLNSV	SYQFLSEQ	455	
sp Q9LVN7 DPOD1_ARATH	NSRVRVRDSTFSSRQ	QGIRESKETTIEGR	RFQFDLIQAIHRDHKLSSY	SLNSVSAHFLSEQ	479	
sp Q9LRE6 DPOD1_ORYSJ	NSRVRVRD	TFSSRQYGMRESKD	VAEGRVQFDLQAMQR	DYKLSSY	SLNSVSAHFLGEQ	489
sp P46588 DPOD_CANAL	RVKQEIKDAFVSSRAYG	TRENKVN	IDGRMQLDDQQF	REYKLRSYTLNSV	SAHFLGEQ	419
sp P97283 DPOD1_MESAU	GLRSNIRDSSFQSRTGRR	DSKV	VMSMGRVQMDMLQVLLRE	HLKLRSYTLNAV	SFHFLGEQ	481
sp O48901 DPOD1_SOYBN	NSRVRVRKDTTFSSRQY	GIRESKETEVAVEGRVTF	DLLQVMQRDYKLSSY	SLNSVSSHFLSEQ	446	
:	:	:*. * : .	* . **	:*.:** ***:*** *. .*** **		

sp P30316 DPOD_SCHPO	KEDVHYSIITD	LQNGTADS	RRRLAIYCLKDAYLPQRLMD	KLMCFVN	YTEEMARVTGVPE	NF	529		
sp P15436 DPOD_YEAST	KEDVHYSIIS	DLQNGDSE	TRRRRLLAVYCLKDAYLP	PLRLMEKL	MALVN	YTEMARVTGVPE	SY	550	
sp P28339 DPOD1_BOVIN	KEDVQHSIIT	DLQNGND	QTRRRRLLAVYCLKDAFLP	LLERLMLV	VNAME	EMARVTGVPL	GY	544	
sp P28340 DPOD1_HUMAN	KEDVQHSIIT	DLQNGND	QTRRRRLLAVYCLKDAYLP	PLRLLLERLMLV	VNAV	EMARVTGVPL	SY	545	
sp O54747 DPOD1_RAT	KEDVQHSIIT	DLQNGNE	QTRRRRLLAVYCLKDAFLP	LLERLMLV	VNNV	EMARVTGVPL	GY	541	
sp P30315 DPOD1_PLAFK	KEDVHYSIM	DLQNEPSRKR	IA	YCIKDGVLPLRL	IDLKLLFYNY	YVEMARVTGTP	FVY	539	
sp P54358 DPOD1_DROME	KEDVHHSIIT	DLQNGD	QTRRRRLLAMYCLKD	AYLPPLR	LLERLML	MAIV	VNYMEMARVTGVPL	ES	529
sp P52431 DPOD1_MOUSE	KEDVQHSIIT	DLQNGNE	QTRRRRLLAVYCLKDAFLP	LLERLMLV	VNNV	EMARVTGVPL	GY	543	
sp P90829 DPOD1_CAEEL	KEDVEHNII	IFDLQRG	DEQTRRRRLLAQYCLKD	AYLPPLR	LLERLMLKMSI	AINYI	EMARVTGVPN	NF	515
sp Q9LVN7 DPOD1_ARATH	KEDVHHSIIT	DLQNGNAE	TRRRRLLAVYCLKD	AYLPQR	LLDKLMFI	YNYV	EMARVTGVPI	SF	539
sp Q9LRE6 DPOD1_ORYSJ	KEDVHHSIIS	DLQNGNAE	TRRRRLLAVYCLKD	AYLPQR	LLDKLMFI	YNYV	EMARVTGVPI	SF	549
sp P46588 DPOD_CANAL	KEDVQHSIIT	DLQNGT	KETR	TRRRRLLAVYCLKDAFLP	LLERLMLC	LNV	YTEMARVTGVPF	SY	479
sp P97283 DPOD1_MESAU	KEDVQHSIIT	DLQNGNE	QTRRRRLLAVYCLR	DAFLP	LLERLMLV	VNNV	EMARVTGVPL	GY	541
sp O48901 DPOD1_SOYBN	KEDVHHSIIS	DLQNGNAE	TRRRRLLAVYCLKD	AYLPQR	LLLDKLMFI	YNYV	EMARVTGVPI	SF	506
****:..:*	***	:*: *:*	***:*	***:***:*	: * . *: ***	***:***:*			

sp P30316 DPOD_SCHPO	LLARGQQIKVISQLFRKALQHDLVVPNIRVNGT-DEQYE	GATVIEPIKGYYDTPIA	TLDL	588
sp P15436 DPOD_YEAST	LLARGQQIKVVSQQLFRKCLEIDTVIPNMQSQAS-DDQYE	GATVIEPIRGYYDVPIA	TLDL	609
sp P28339 DPOD1_BOVIN	LLSRGQQVKVVSQQLRQAMRGQLLMPVVKTEGG--EDYI	GATVIEPLKGYYDVPIA	TLDL	602
sp P28340 DPOD1_HUMAN	LLSRGQQVKVVSQQLRQAMHEGLLMPVVKTEGG--EDYI	GATVIEPLKGYYDVPIA	TLDL	603
sp O54747 DPOD1_RAT	LLSRGQQVKVVSQQLRQAMREGGLLMPVVKTEGG--EDYI	GATVIEPLKGYYDVPIA	TLDL	599
sp P30315 DPOD1_PLAFK	LLTRGQQIKVTSQLYRKCKELNYVIPSTYMKVNTNEKYE	GATVIEPIKGYYIEPI	TLDL	599
sp P54358 DPOD1_DROME	LLTRGQQIKVLSQQLRKAKTKGFIMPSTQSGS-DEQYE	GATVIEPKRGYYADPI	TLDL	588
sp P52431 DPOD1_MOUSE	LLTRGQQVKVVSQQLRQAMRGQLLMPVVKTEGG--EDYI	GATVIEPLKGYYDVPIA	TLDL	601
sp P90829 DPOD1_CAEEL	LLTKGQQIKILSMSMLRRCKQNFFLPVIEANSGDGEGYF	GATVIDPIRGFYNEPI	TLDL	575
sp Q9LVN7 DPOD1_ARATH	LLARGQSIKVLSQLLRKGKQKNLVLVPNQAGSEQGTFF	GATVLEARTGFYEKPI	TLDL	599
sp Q9LRE6 DPOD1_ORYSJ	LLSRGQSIKVLSQLLRKAQKNLVIPIQAGSEQGTFF	GATVLEARAGFYEKPI	TLDL	609
sp P46588 DPOD_CANAL	LLSRGQSIKVISQLFRKLCQEDIVIPNLKSEGS-NEEYF	GATVIEPERGYDVPIA	TLDL	538
sp P97283 DPOD1_MESAU	LLSRGQSIKVVSQQLRQAMRGQLLMPVVKTEGG--EDYI	GATVIEPLKGYYDVPIA	TLDL	599
sp O48901 DPOD1_SOYBN	LLSRGQSIKVLSQLLRARQKNLVIPIQAGSEQGTFF	GATVLEARAGFYEKPI	TLDL	566
	: : * : * : . : * :	: ****: : . *: * : ***: ****		
sp P30316 DPOD_SCHPO	SSLYPSIMQAHNLCYTTILDNSTAELL----KLKDQDV-DYSVTPNGDYFVKPHVRKGLLP			643
sp P15436 DPOD_YEAST	SSLYPSIMMAHNLCYTTLCKNATVERL---NLKIDE-DYVTPNGDYFVTTKRRGILP			564
sp P28339 DPOD1_BOVIN	SSLYPSIMMAHNLCYTTLRPGAQAKL---GLTED--QFIKPTGDEFVKASVRKGLLP			656
sp P28340 DPOD1_HUMAN	SSLYPSIMMAHNLCYTTLRPGAQAKL---GLTED--QFIRKPTGDEFVKTSVRKGLLP			657
sp O54747 DPOD1_RAT	SSLYPSIMMAHNLCYTTLRPGAQAKL---GLKPD--EFIGKPTGDEFVKASVRKGLLP			553
sp P30315 DPOD1_PLAFK	ASLYPSIMIAHNLCYSTLIKSNH-EVS----DLQNNDDITTIQGKNNLKFVKKNVKKGILP			554
sp P54358 DPOD1_DROME	ASLYPSIMMAHNLCYTTVLGGTREKLRQQENLQDD--QVERTPANNYFVKSEVRGLLP			646
sp P52431 DPOD1_MOUSE	SSLYPSIMMAHNLCYTTLRPGAQAKL---GLKPD--EFIGKPTGDEFVKSSVRKGLLP			555
sp P90829 DPOD1_CAEEL	ASLYPSIMIAHNLCYTTLKSP---Q---GVENE--DYIRTSPSQYFATKSKRRGILP			525
sp Q9LVN7 DPOD1_ARATH	ASLYPSIMMAYNLCYTLPEDVRKL---NLPP--HVTKTPSGETFKQTQKGILP			553
sp Q9LRE6 DPOD1_ORYSJ	ASLYPSIMMAYNLCYTLPEDARKL---NLPP--SVNKTPSGETFKPDKVQKGILP			563
sp P46588 DPOD_CANAL	SSLYPSIMMAHNLCYTTLNKNNSIKAF---GLTED--DYTKTPTGDEFVKSSVRKGLLP			592
sp P97283 DPOD1_MESAU	SSLYPSIMMAHNLCYTLLRPGAQAKL---GLKPD--EFIGKPTGDEFVKSSVRKGLLP			553
sp O48901 DPOD1_SOYBN	ASLYPSIMMAYNLCYTLPEDARKL---NIPPE--SVNRTPSGETFKSNLQKGILP			620
	*****: *:****: * :	: : . * . : :*:***		
sp P30316 DPOD_SCHPO	IILADLLNARKAKADLKETDPFKKAVLDRQLALKVSANSVYGTGATV-TN-GRLPCIAI			702
sp P15436 DPOD_YEAST	IILDELISARKRAKKDLRDEKDPFKRDVLNGRQLALKISANSVYGTGATV-GKLPCIAI			723
sp P28339 DPOD1_BOVIN	QILENLISARKRAKAEELAKETDPPLRQVLDRQLALKVSANSVYGTGATV-QV-GRLPCLEI			715
sp P28340 DPOD1_HUMAN	QILENLISARKRAKAEELAKETDPPLRQVLDRQLALKVSANSVYGTGATV-QV-GKLPCLEI			716
sp O54747 DPOD1_RAT	QILENLISARKRAKAELAQETDPPLRQVLDRQLALKVSANSVYGTGATV-QV-GKLPCLEI			712
sp P30315 DPOD1_PLAFK	LIVEELIARKKVKLLIKNEKNNTKMLNGRQLALKISANSVYGTGASSGGQQLPCLEV			714
sp P54358 DPOD1_DROME	EILEELIARKRAKNDLKVKETDPFKRKVLDRQLALKISANSVYGTGATV-QV-GKLPCLEI			705
sp P52431 DPOD1_MOUSE	QILENLISARKRAKAELAQETDPPLRQVLDRQLALKVSANSVYGTGATV-QV-GKLPCLEI			714
sp P90829 DPOD1_CAEEL	EILEDILIAARKRAKNDMKNEKDEFKRMVYNGRQLALKISANSVYGTGATV-GKLPCLEI			684
sp Q9LVN7 DPOD1_ARATH	EILEELITARKRAKADLKEAKDPLEKAVLDRQLALKISANSVYGTGATV-GQLPCLEI			712
sp Q9LRE6 DPOD1_ORYSJ	EILEELITARKRAKADLKEAKDPFERAVLDRQLALKISANSVYGTGATV-GQLPCLEI			722
sp P46588 DPOD_CANAL	TILDELLTARKKAKADLKETDPFKKDVLNGRQLALKISANSVYGTGATV-GKLPCIAI			651
sp P97283 DPOD1_MESAU	QILENLISARKRAKAELAQETDPPLRQVLDRQLALKVSANSVYGTGAEV-GKLPCLEI			712
sp O48901 DPOD1_SOYBN	EILEELITARKRAKADLKEAKDPLEKAVLDRQLALKISANSVYGTGATI-GQLPCLEI			679
	*: . : ***: * : . : : * : ****: *:****:***	*:****: :		

sp P30316 DPOD_SCHPO	SSSVTSYGRQMIEKTKDVVEKRYR---IENGYSHDAVVIYGDTDSVMVKFGVKTLPEAMK	759
sp P15436 DPOD_YEAST	SSSVTAYGRTMILKTKTAVQEKYC---IKNGYKHDAVVYGDTSVMVKFGTTDLKEAMD	780
sp P28339 DPOD1_BOVIN	SQSVTGFGRQMIEKTKQLVETKYT---VENGSTSAKVVYGDTSVMCRFGVSSVAEAMA	772
sp P28340 DPOD1_HUMAN	SQSVTGFGRQMIEKTKQLVESKYT---VENGSTSAKVVYGDTSVMCRFGVSSVAEAMA	773
sp O54747 DPOD1_RAT	SQSVTGFGRQMIEKTKQLVETKYT---LENGYDANAKVVYGDTSVMCRFGVSSVAEAMS	769
sp P30315 DPOD1_PLAFK	AVSITTLGRSMIEKTKERVESFYC---KSNGYEHNSTVIYGDTDSVMVKFGTNNIEEAMT	771
sp P54358 DPOD1_DROME	SGSVTAYGRTMIEMTKNEVESHT---QANGYENNNAVIYGDTDSVMVNFGVKTLERSME	762
sp P52431 DPOD1_MOUSE	SQSVTGFGRQMIEKTKQLVESKYT---VENGYDANAKVVYGDTSVMCRFGVSSVAEAMS	771
sp P90829 DPOD1_CAEEL	SQSVTAFGRKMIDMTKLEVERIYKKGALDGKCPADAKVIYGDTDSVMVKFGVETVAQAME	744
sp Q9LVN7 DPOD1_ARATH	SSSVTSYGRQMIEQTKKLVEDKFT---TLGGYQYNAEVVIYGDTDSVMVQFGVSDEAAMT	769
sp Q9LRE6 DPOD1_ORYSJ	SSSVTSYGRQMIEHTKKLVEDKFT---TLGGYEHNAEVVIYGDTDSVMVQFGVSTVEDAMK	779
sp P46588 DPOD_CANAL	SSSVTAFCREMIEKTKNEVQEYYS---KKNGHHPYDAKVIYGDTDSVMVKFGYQDLETCKM	708
sp P97283 DPOD1_MESAU	SQSVTGFGRQMIEKTKQLVESKYT---LENGYNANAKVVYGDTSVMCRFGVSSVAEAMS	769
sp O48901 DPOD1_SOYBN	SSSVTSYGRQMIEHTKKLVEDKFT---TLNGYEHNAEVVIYGDTDSVMVQFGVSAVEEAMN	736
	:	***** .** :

sp P30316 DPOD_SCHPO	LGEEAANYVSDQFPNPIKLEFEKVPYPYLLISKKRYAGLFW-TRTDTYDKMD\$KGIELTVR	818
sp P15436 DPOD_YEAST	LGTEAAKYVSTLFKHPINLEFEKAYFPYLLISKKRYAGLFW-TNPDKFDKLDOKGLASVR	839
sp P28339 DPOD1_BOVIN	LGREAADWVSGHFPSPIRLEFEKVYPYLLISKKRYAGLFW-FSSRPDAHDRMDCKGLEAVR	832
sp P28340 DPOD1_HUMAN	LGREAADWVSGHFPSPIRLEFEKVYPYLLISKKRYAGLFW-FSSRSRDAHDRMDCKGLEAVR	833
sp O54747 DPOD1_RAT	LGREAANWVSSHFPSPPIRLEFEKVYPYLLISKKRYAGLFW-FSSRSRDAHDRMDCKGLEAVR	829
sp P30315 DPOD1_PLAFK	LGKDAAERISKEFLSPIKLEFEKVCYCPYLLLNKKRYAGLLY-TNPNKHDKMDCKGIELTVR	830
sp P54358 DPOD1_DROME	LGREAABLVSSKFVHPPIKLEFEKVYPYLLLNKKRYAGLWF-TRPDTYDKMDCKGIELTVR	821
sp P52431 DPOD1_MOUSE	LGREAANWVSSHFPSPPIKLEFEKVCYCPYLLLNKKRYAGLWF-FSSRSRDAHDRMDCKGLEAVR	831
sp P90829 DPOD1_CAEEL	IGLDAAKEVSKIFTTPPIKLEFEKVPYSYPLLNKKRYAGLWF-TKPDVHDKMDCKGLEAVR	803
sp Q9LVN7 DPOD1_ARATH	LGREAAEHISGTFIKPIKLEFEKVPYSYPLLNKKRYAGLWF-TNPQQFDKMDTCKGIELTVR	828
sp Q9LRE6 DPOD1_ORYSJ	LGREADDYISGTFIKPIKLEFEKIYFPYLLISKKRYAGLWF-TNPEKFDFKMDTCKGIELTVR	838
sp P46588 DPOD_CANAL	LGEEAANYVSTKFKNPIKLEFEKVPYSYPLLNKKRYAGLWF-TRPEKFDFKMDTCKGIELTVR	767
sp P97283 DPOD1_MESAU	LGREAANWVSSHFPSPPIKLEFEKVCYFPYLLISKKRYAGLFW-FSSQPDTHDRMDCKGLEAVR	829
sp O48901 DPOD1_SOYBN	LGREAAEHISGTFIKPIKLEFEKVCYFPYLLISKKRYAGLFW-TKPDNFDFKMDTCKGIELTVR	795
	:	*** . : * * . * . * . * . * . * . * . * . * . * . * . * . * . * . * . * . * .

sp P30316 DPOD_SCHPO	RDNCPILVSYVIDTALRKMLIDQDVEGAQLFTKKVISDLLQNIDMSQLVITKALSK--TD	876
sp P15436 DPOD_YEAST	RDSCSLVSIVMNKVLKKILLERNDGALFVRETINDILHNRVDISKLIISKTLA--PN	896
sp P28339 DPOD1_BOVIN	RDNCPILVNLVTASLRKLIDRDPGAVAHQAQDVISIDLNCRIDISQLVITKELTRAAD	892
sp P28340 DPOD1_HUMAN	RDNCPILVNLVTASLRLLIDRDPGEGAVAHQAQDVISIDLNCRIDISQLVITKELTRAASD	893
sp O54747 DPOD1_RAT	RDNCPILVNLVTSSLRRLIDRDPGAVAHAKDVISIDLNCRIDISQLVITKELTRAAD	889
sp P30315 DPOD1_PLAFK	RDFCILIQQMETVLNKLLIEKNLNSAIEYTKSKIKELLTNNIDMSLLVVTKSLGK--TD	888
sp P54358 DPOD1_DROME	RDNSPVLVANLMNSCLQKLLIERDPDGAVAVYVKQVIADLLCNRIDISHLVITKELAK--TD	879
sp P52431 DPOD1_MOUSE	RDNCPILVNLVTSSLRRLIDRDPGAVAHAKDVISIDLNCRIDISQLVITKELTRAAD	891
sp P90829 DPOD1_CAEEL	RDNCPILVAKVLGVCLEKLLIERDQQSALDFAKRTISIDLNCRIDISLLIIISKELTKSGDK	863
sp Q9LVN7 DPOD1_ARATH	RDNCLLVKNLVTECLHKILVDRDPGAVQYVKNNTISDLLMNMRIDLSSLVITKGLTKTGDD	888
sp Q9LRE6 DPOD1_ORYSJ	RDNCLLVQNVITKVLEFILEERDVPKAQRFKQTIADLLQNRIDLSQLVITKAYSK--HD	825
sp P46588 DPOD_CANAL	RDNCPILVNLVTSSLRRLIDRDPGAVAHAKDVISIDLNCRIDISQLVITKELTRAAD	889
sp P97283 DPOD1_MESAU	RDNCLLVKNLVNDCLHKILIDRDIPIPGAVQYVKNASIDLMMNRMDLSSLVITKGLTKTGDD	855
sp O48901 DPOD1_SOYBN	*** . * : :: * . * : :: * . : * . * : * . * : * . * : * .	.

sp P30316 DPOD_SCHPO	YAAKMAHVELAERMRKR DAGSAPA <u>I</u> GDRVAYVIIKGAGQGDQFYMRSEDPIYVLENNIPID	936													
sp P15436 DPOD_YEAST	YTNPQPHAVLAERMRKRRE-GVGPNVGDRV <u>D</u> YVIIGG--NDKLYNRAEDPLFVLENNIQVD	953													
sp P28339 DPOD1_BOVIN	YAGKQAHVELAERMRKR DPGSAPS <u>L</u> GDRV <u>P</u> YVIISAAKGVAA YMKS EDP L FVLEHSLPID	952													
sp P28340 DPOD1_HUMAN	YAGKQAHVELAERMRKR DPGSAPS <u>L</u> GDRV <u>P</u> YVIISAAKGVAA YMKS EDP L FVLEHSLPID	953													
sp O54747 DPOD1_RAT	YAGKQAHVELAERMRKR DPGSAPN <u>L</u> GDRV <u>P</u> YVIIGAAKGVAA YMKS EDP L FVLEHSLPID	949													
sp P30315 DPOD1_PLAFK	YETRLPHVELAKKLQQRDSATAPN <u>V</u> GDRVSYIIVKGVKGQAQYERAEDPL YVLDNNLAID	948													
sp P54358 DPOD1_DROME	YAAKQAHVELAAMKKR DPGTAPK <u>L</u> GDRV <u>P</u> YVICAACAKNTPAYQKAEDPL YVLENSLPID	939													
sp P52431 DPOD1_MOUSE	YAGKQAHVELAERMRKR DPGSAPS <u>L</u> GDRV <u>P</u> YVIIGAAKGVAA YMKS EDP L FVLEHSLPID	951													
sp P90829 DPOD1_CAEEL	YQAKQAHVELAARMKKR DAGSAPR <u>L</u> GDRV <u>P</u> YVFVAAKNVPAYERAEDPTFV LQNNIPLD	923													
sp Q9LVN7 DPOD1_ARATH	YEVKSAHGLEAERMRKRDAA TAPN <u>V</u> GDRV <u>P</u> YVIKA AKGAKAYERSEDP IYV LQNNIPLD	948													
sp Q9LR66 DPOD1_ORYSJ	YAVKAAHVELAERMRKRDAA TAPV <u>G</u> DRV <u>P</u> YVIKA AKGAKAYERSEDP IYV LQNNIPLD	958													
sp P46588 DPOD_CANAL	YSAKQAHVELAERMRKR DPGSAP <u>T</u> L GDRV <u>P</u> YVIIGAAKGVAA YMKS EDP L FVLEHSLPID	884													
sp P97283 DPOD1_MESAU	YAGKQAHVELAERMRKR DPGSAPS <u>L</u> GDRV <u>P</u> YVIIGAAKGVAA YMKS EDP L FVLEHSLPID	949													
sp O48901 DPOD1_SOYBN	YEVKAAHVELAERMRKRDAA TAPN <u>V</u> GDRV <u>P</u> YVIKA AKGAKAYERSEDP IYV LQNNIPLD	915													
	*	.	*	**	:	:	***	:	***	:	***	:	***	:	*

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.

Fig. 2K Epsilon polymerases (Eukaryotes)

gi 285814403 tpg DAA10297.1	AQQLLNASKIDDIDSMMGFERYVPPQYNG-RFDAKDIDQI PGRVGWL TNM	82													
gi 118841 sp P21951.1 DPOE_YEA	AQQLLNASKIDDIDSMMGFERYVPPQYNG-RFDAKDIDQI PGRVGWL TNM	82													
gi 259485986 tpe CBF83469.1	SLRSSEATSQDEKFEAIRLANSIDESLGFPFREAGEK-----RVGWL INM	76													
gi 194214443 ref XP_001915746.	QLGRRDDGPSSSSVALKRLERSQWTDKMDLRLFGFERLREPGEKT GWL INM	100													
Q07864-1	-----RDDGATSSVSALKRLERSQWTDKMDLRLFGFERLKEPGEKT GWL INM	66													
	:	.	..	:	.	.	*	**	:	***	**				
gi 285814403 tpg DAA10297.1	HATLVSQETLSSGSNGGGNSNDGERVTTNQGISGVDFYFLDEEG GSFKST	132													
gi 118841 sp P21951.1 DPOE_YEA	HATLVSQETLSSGSNGGGNSNDGERVTTNQGISGVDFYFLDEEG GSFKST	132													
gi 259485986 tpe CBF83469.1	HSTSIEDPNPGGR-----AGV DYYFLDDDGGSF KAT	108													
gi 194214443 ref XP_001915746.	HPTEVLDEDKRLVS-----AVDYYFIQDDGSRFKVA	131													
Q07864-1	HPTEILDEDKRLGS-----AVDYYFIQDDGSRFKVA	97													
	*	.	*	:	..	:	..	*	**	:	***	**			
gi 285814403 tpg DAA10297.1	VVYDPYFFIA C NDES RVNDVEELVKKYLESLCKSLQIIRKEDLTMDN HLL	182													
gi 118841 sp P21951.1 DPOE_YEA	VVYDPYFFIA C NDES RVNDVEELVKKYLESLCKSLQIIRKEDLTMDN HLL	182													
gi 259485986 tpe CBF83469.1	VEYDPYFLIA VKTG-HEAEVEEWCRRMFEGLIK KIKR VV KEDLKL PN HLL	157													
gi 194214443 ref XP_001915746.	LPYKPYFYI ATRKG-CEREVSSFLSKKFQGKIAK VETVPEKEDL DLPN HLV	180													
Q07864-1	LPYKPYFYI ATRKG-CEREVSSFLSKKFQGKIAK VETVPEKEDL DLPN HLV	146													
	:	*	***	**	.	:	* ..	:	..	:	..	:	***	:	***
gi 285814403 tpg DAA10297.1	-----NVAANGSEKVDAKHLIEDIREYDVPYHVRV S IDK DIRVGK WY KV	267													
gi 118841 sp P21951.1 DPOE_YEA	-----NVAANGSEKVDAKHLIEDIREYDVPYHVRV S IDK DIRVGK WY KV	267													
gi 259485986 tpe CBF83469.1	-----ARPNGTTNASDFI DIREYDVPYHVRV AIDK DIRIGK WY TV	232													
gi 194214443 ref XP_001915746.	GGSVIIDEEETSKMADQLDNIVDMREYDVPYHIRLSIDLKIHVAHWYNV	280													
Q07864-1	RGGVITDEEETSKIADQLDNIVDMREYDVPYHIRLSIDLKIHVAHWYNV	246													
	.	..	:	*	*****	:	***	.	**	:	***	*			

gi 285814403 tpg DAA10297.1 gi 118841 sp P21951.1 DPOE_YEA gi 259485986 tpe CBF83469.1 gi 194214443 ref XP_001915746. Q07864-1	TQQG-----FIEDTRKIAFAADPVVMAFDIETTKPPLKFPSAVDQIMMI 311 TQQG-----FIEDTRKIAFAADPVVMAFDIETTKPPLKFPSAVDQIMMI 311 EATHGII--SLTCLEERLTRAAPVVLAFDIETTKLPLKFPSVIDQIMMI 280 RYRGNNAFPVEIARRDDLVERPDPVVLAFLDIETTKLPLKFDAETTDQIMMI 330 RYRGNNAFPVEITRRDDLVERPDPVVLAFLDIETTKLPLKFDAETTDQIMMI 296 : : .*****:*****:*****:*****:*****:*****:
gi 285814403 tpg DAA10297.1 gi 118841 sp P21951.1 DPOE_YEA gi 259485986 tpe CBF83469.1 gi 194214443 ref XP_001915746. Q07864-1	SYMIDGEGLITNREIISEDIEDFEYTPKPEYPGFFTIFNENDEVALLQR 361 SYMIDGEGLITNREIISEDIEDFEYTPKPEYPGFFTIFNENDEVALLQR 361 SYMIDGQGLITNREIVSEDIDDFEYTPKPEYSGPFMIFNENPNERAVIER 330 SYMIDGQGYLITNREIVSEDIEFETPKPEYEGLFCVFNEPDEVHLIQR 380 SYMIDGQGYLITNREIVSEDIEFETPKPEYEGLFCVFNEPDEAHLIQR 346 *****:*****:*****:*****:*****:*****:*****:*****:*****:*****:
gi 285814403 tpg DAA10297.1 gi 118841 sp P21951.1 DPOE_YEA gi 259485986 tpe CBF83469.1 gi 194214443 ref XP_001915746. Q07864-1	FFEHHRDVRPTVISTFNGDFFDWPFHNRSKIHGLDMFDEIGFAPDAEGER 411 FFEHHRDVRPTVISTFNGDFFDWPFHNRSKIHGLDMFDEIGFAPDAEGER 411 FFEHRIKEAKPTVIATYNGDFFDWPFVEARASVLGIDMYKEIGFRKNSEDI 380 WFEHVQETKPTIMVTYNGDFFDWPFVEARAAVHGLSMQEIGFQKDNQGE 430 WFEHVQETKPTIMVTYNGDFFDWPFVEARAAVHGLSMQEIGFQKDSQGE 396 :*****:*****:*****:*****:*****:*****:*****:*****:*****:
gi 285814403 tpg DAA10297.1 gi 118841 sp P21951.1 DPOE_YEA gi 259485986 tpe CBF83469.1 gi 194214443 ref XP_001915746. Q07864-1	YKSSYCSHMDCFRWVKRDSDYLPQGSQGLKAVTQSKLGYNPIELDPELMTP 461 YKSSYCSHMDCFRWVKRDSDYLPQGSQGLKAVTQSKLGYNPIELDPELMTP 461 YQSDHCAHMDCFAWVNDRDSYLPQGSRGGLKAVTVAKLGYDDELDPPELMTP 430 YKAPQCIHMDCLRWVKRDSDYLPVGSHNLKAAAKAKLGYPVELDPEDMCR 480 YKAPQCIHMDCLRWVKRDSDYLPVGSHNLKAAAKAKLGYPVELDPEDMCR 446 *: * *****: *:*****: *:*****: *:*****: *:*****: *:*****:
gi 285814403 tpg DAA10297.1 gi 118841 sp P21951.1 DPOE_YEA gi 259485986 tpe CBF83469.1 gi 194214443 ref XP_001915746. Q07864-1	YAFEKPKQLSEYSVSDAVATYYLYMKVYHPFIFSLCTIIPLNPDETLRKG 511 YAFEKPKQLSEYSVSDAVATYYLYMKVYHPFIFSLCTIIPLNPDETLRKG 511 YASERPQTIAEYSVSDAVATYYLYMKVYHPFIFSLCTIIPLNPDETLRKG 480 MATEQPQTLATYSVSDAVATYYLYMKVYHPFIFALCTIIPMEPDEVLRKG 530 MATEQPQTLATYSVSDAVATYYLYMKVYHPFIFALCTIIPMEPDEVLRKG 496 *: *: *: *:*****:*****:*****:*****:*****:*****:*****:
gi 285814403 tpg DAA10297.1 gi 118841 sp P21951.1 DPOE_YEA gi 259485986 tpe CBF83469.1 gi 194214443 ref XP_001915746. Q07864-1	TGTLCMMLMVQAYQHNILLPNKHTDPIERFYDG-HILLESETYVGHHVES 560 TGTLCMMLMVQAYQHNILLPNKHTDPIERFYDG-HILLESETYVGHHVES 560 TGTLCMMLMVQAYKGNIILVLPNKHKDPEAFYEG-HILLESETYVGHHVES 529 SGTLCEALLMVQAFHANIIFPNKQEQUEFNKLTDGGHVLDSETYVGHHVEA 580 SGTLCEALLMVQAFHANIIFPNKQEQUEFNKLTDGGHVLDSETYVGHHVEA 546 :*****:*****:*****:*****:*****:*****:*****:
gi 285814403 tpg DAA10297.1 gi 118841 sp P21951.1 DPOE_YEA gi 259485986 tpe CBF83469.1 gi 194214443 ref XP_001915746. Q07864-1	LEAGVFRSDLKNEFKIDPSAIDELLOELPEALKFSVEVENKSSVDKVTNF 610 LEAGVFRSDLKNEFKIDPSAIDELLOELPEALKFSVEVENKSSVDKVTNF 610 IEAGVFRSDLIPVPFNIDPTAVDELLRLDAALKFSIEVEEKKSSLDDVTNY 579 LESGVFRSDIPCRFRMNPAAFDPLLQRLRVEKTLRHAIEEEEKVPVEQVTNF 630 LESGVFRSDIPCRFRMNPAAFDPLLQRLRVEKTLRHAIEEEEKVPVEQVTNF 596 *:*****: *:*****: *:*****: *:*****: *:*****:

gi|285814403|tpg|DAA10297.1|
 gi|118841|sp|P21951.1|DPOE_YEA
 gi|259485986|tpe|CBF83469.1|
 gi|194214443|ref|XP_001915746.
 Q07864-1

EEIKNQITQKLLELKENNIRNELPLIYHVDVASMYPNIMTTNRLQPDSIK 660
 EEIKNQITQKLLELKENNIRNELPLIYHVDVASMYPNIMTTNRLQPDSIK 660
 EEVKGQIAKLLTDLRENPHRNEVPFIYHLDVASMYPNIMTTNRLQPDSIK 629
 QEVCDQIKTKLTSKLDVDPNRIECPLIYHLDVGAMYPNIILTNRLQPSAMV 680
 EEVCDIEIKSKLASLKDVPNSRIECPLIYHLDVGAMYPNIILTNRLQPSAMV 646
 * : . * . * . : * * : * : * : * : * : * : * : * : * : * :

gi|285814403|tpg|DAA10297.1|
 gi|118841|sp|P21951.1|DPOE_YEA
 gi|259485986|tpe|CBF83469.1|
 gi|194214443|ref|XP_001915746.
 Q07864-1

AERDQASCDFNRPCKTCARKLKWAWRGEFFPSKMDEYNMIKRALQNETFP 710
 AERDQASCDFNRPCKTCARKLKWAWRGEFFPSKMDEYNMIKRALQNETFP 710
 QESNCACDFNRPCKTCARKLKWAWRGEFFPSKMDEYNMIKRALQNETFP 679
 DEATCAACDFNPKPGANCQRKMAWQWRGEFMPASRSEYHRIHQLESEKFP 730
 DEATCAACDFNPKPGANCQRKMAWQWRGEFMPASRSEYHRIHQLESEKFP 696
 * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :

gi|285814403|tpg|DAA10297.1|
 gi|118841|sp|P21951.1|DPOE_YEA
 gi|259485986|tpe|CBF83469.1|
 gi|194214443|ref|XP_001915746.
 Q07864-1

NKNKFSKKVLTFDELSYADQVIHIKKRLTEYSRKVYHRVKVSEIVEREA 760
 NKNKFSKKVLTFDELSYADQVIHIKKRLTEYSRKVYHRVKVSEIVEREA 760
 GRTKSP--MRAFTELSEAEEQAAIVKKRQLDYSKKIYHKIHDSTMVREA 727
 PLTPEGP--ARAFHELSREEQAKYEKRRLADYCRKAYKKIHKVTKVEERLT 778
 PLFPEGP--ARAFHELSREEQAKYEKRRLADYCRKAYKKIHKVTKVEERLT 744
 : * : * : * : * : * : * : * : * : * : * : * : * : * : * :

gi|285814403|tpg|DAA10297.1|
 gi|118841|sp|P21951.1|DPOE_YEA
 gi|259485986|tpe|CBF83469.1|
 gi|194214443|ref|XP_001915746.
 Q07864-1

IVCQRENPFYVDTVKSFRDRRYEFKGAKTWKGNLSKIDPSDK--HARDE 808
 IVCQRENPFYVDTVKSFRDRRYEFKGAKTWKGNLSKIDPSDK--HARDE 808
 TICQRENPFYVDTVRSFRDRRYEFKGKQKVWKGKTESLKGSSAPAAEIEE 777
 TICQRENSFYVDTVRAFRDRRYEFKGHLKVWKKLSSADEVGD-AAEVKR 827
 TICQRENSFYVDTVRAFRDRRYEFKGHLKVWKKLSSADEVGD-AAEVKR 793
 : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :

gi|285814403|tpg|DAA10297.1|
 gi|118841|sp|P21951.1|DPOE_YEA
 gi|259485986|tpe|CBF83469.1|
 gi|194214443|ref|XP_001915746.
 Q07864-1

AKKMIIVLYDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGITCLTGAIII 858
 AKKMIIVLYDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGITCLTGAIII 858
 AKKMIIVLYDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGITCLTGAIII 827
 CRNMEVLYDSLQLAHKCILNSFYGYVMRKGSRWYSMEMAGIVCFTGANII 877
 CKNMEVLYDSLQLAHKCILNSFYGYVMRKGSRWYSMEMAGIVCFTGANII 843
 : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :

gi|285814403|tpg|DAA10297.1|
 gi|118841|sp|P21951.1|DPOE_YEA
 gi|259485986|tpe|CBF83469.1|
 gi|194214443|ref|XP_001915746.
 Q07864-1

QMALARVERVGRPLELTDGIWCILPKSFPETYFFTLENGKK--LYLSYP 906
 QMALARVERVGRPLELTDGIWCILPKSFPETYFFTLENGKK--LYLSYP 906
 QMARELVERIGRPLELTDGIWCMLPGTFPENFSFTLKNGKK--LGISYP 875
 TQARELIEQIGRPLELTDGIWCVLPNSPENFVVKTTSVKPKVTISYP 927
 TQARELIEQIGRPLELTDGIWCVLPNSPENFVFKTTNVKPKVTISYP 893
 * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :

gi|285814403|tpg|DAA10297.1|
 gi|118841|sp|P21951.1|DPOE_YEA
 gi|259485986|tpe|CBF83469.1|
 gi|194214443|ref|XP_001915746.
 Q07864-1

C\$MLNIVRHQKFTNHQYQELKDPLNYIYETHSENTIFFEVDPYKAMILP 956
 C\$MLNIVRHQKFTNHQYQELKDPLNYIYETHSENTIFFEVDPYKAMILP 956
 CVMLNHLVHGSYTNHQYQSLANPATFHYYETHSENSIFFEVDPYRAMILP 925
 GAMLNIMVKEGFTNDQYQELAEPSSLTYVTRSENSIFFEVDPYLAMILP 977
 GAMLNIMVKEGFTNDQYQELAEPSSLTYVTRSENSIFFEVDPYLAMILP 943
 * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :

gi 285814403 tpg DAA10297.1 gi 118841 sp P21951.1 DPOE_YEA gi 259485986 tpe CBF83469.1 gi 194214443 ref XP_001915746. Q07864-1	SSKEEGKGIIKKRYAVFNEGDGLAELKGFELKRRGELQLIKNFQS DIFKVF 1006 SSKEEGKGIIKKRYAVFNEGDGLAELKGFELKRRGELQLIKNFQS DIFKVF 1006 TSKEEDKNIKKRYAVFNDDGSLAELKGFEVKRRGELKLKIFQT QIFKFF 975 ASKEEGKKIJKKRYAVFNEGDGLAELKGFEVKRRGELQLIKIFQS SVFEAF 1027 ASKEEGKKIJKKRYAVFNEGDGLAELKGFEVKRRGELQLIKIFQS SVFEAF 993 :*****.* :*****:*****:*****:*****:*** *: .: *:
gi 285814403 tpg DAA10297.1 gi 118841 sp P21951.1 DPOE_YEA gi 259485986 tpe CBF83469.1 gi 194214443 ref XP_001915746. Q07864-1	LEGDTLEGCYSAVASVCNRWLDVLDSHGLMLEDEDLVS LICENRSMSTKL 1056 LEGDTLEGCYSAVASVCNRWLDVLDSHGLMLEDEDLVS LICENRSMSTKL 1056 LEGTLAETYAAVARVADRWLDVLVEHGAITLEADEELIELISENRSMTKTL 1025 LKGSTLEEVYGSVAKADVWLDVLVSKAANMPDSELFELISENRSMSRKL 1077 LKGSTLEEVYGSVAKADVWLDVLVSKAANMPDSELFELISENRSMSRKL 1043 *: * * .: * * .: * * * .. : * .: * .: * * .: * * .: *
gi 285814403 tpg DAA10297.1 gi 118841 sp P21951.1 DPOE_YEA gi 259485986 tpe CBF83469.1 gi 194214443 ref XP_001915746. Q07864-1	KEYEQKSTSITTARRLGDFLGEDMVKDGLQCKYIIS SKPFNA PVTERA 1106 KEYEGQKSTSITTARRLGDFLGEDMVKDGLQCKYIIS SKPFNA PVTERA 1106 EEYGNQKSTSITTARRLAEFLGEQMVVKDGNCCKYIIS ARPRNT PVTERA 1075 EDYGEQKSTSISTAKRLAEFLGDQMVKDAGLSCRYSRKPEGSPVTERA 1127 EDYGEQKSTSISTAKRLAEFLGDQMVKDAGLSCRYSRKPEGSPVTERA 1093 . : * *****: * : * : * : * : * : * : * : * .: * : * : * :
gi 285814403 tpg DAA10297.1 gi 118841 sp P21951.1 DPOE_YEA gi 259485986 tpe CBF83469.1 gi 194214443 ref XP_001915746. Q07864-1	I PVAIF SADIPIKRSFLRRWT LDPGLELDLDIRTIIDWGYYRERLGSAIQK 1156 I PVAIF SADIPIKRSFLRRWT LDPGLELDLDIRTIIDWGYYRERLGSAIQK 1156 I PVTIIFSAEDSIKRHFLRKWLKD-DPGMDMPSVIDDWYYIERLGSVVKQK 1124 I PLAIFIQAEPVRKHFLRKWLKSSSLQEFDIRTILDWD YYIERLGSAIQK 1177 I PLAIFIQAEPVRKHFLRKWLKSSSLQDFDIRAILDW YYIERLGSAIQK 1143 . : * : * .: * : * : * .: * : * : * : * : * : * : * : * :
gi 285814403 tpg DAA10297.1 gi 118841 sp P21951.1 DPOE_YEA gi 259485986 tpe CBF83469.1 gi 194214443 ref XP_001915746. Q07864-1	IITIPAAALQGVSNPVPVREHPDWIKRKIAKEDFKQTSITKFFSK--TK 1204 IITIPAAALQGVSNPVPVREHPDWIKRKIAKEDFKQTSITKFFSK--TK 1204 IITIPAAALQKIRNPVPRVAHPEWIQRINKQDRFKQVKMFTDMFGK--SE 1172 IITIPAAALQOVKNPVPVPRVKHPDWIHKKLLKEKNDIYKQKKISEFILEGKR 1227 IITIPAAALQOVKNPVPVPRVKHPDWIHKKLLKEKNDVYKQKKISELEFTLEGR 1193 . : * : * : * : * : * : * : * : * : * : * : * : * : * :
gi 285814403 tpg DAA10297.1 gi 118841 sp P21951.1 DPOE_YEA gi 259485986 tpe CBF83469.1 gi 194214443 ref XP_001915746. Q07864-1	VLEVFTTINGKVQNITFHIPKTIYMKFKSQTMLQKIKNCLIEKSSASLP 1384 VLEVFTTINGKVQNITFHIPKTIYMKFKSQTMLQKIKNCLIEKSSASLP 1384 IVRAFVLIDRKIHALTIVPRCVYINLKQDSLDPDVPECEVEKVNHTLP 1359 LFRLWAVISSDLYCIKLNIPRVFYVNQR---VAKAEGPSYRKVNRLVP 1423 LFRLWALVGSDLHCIRLSIPRVFYVNQR---VAKAEGASYRKVNRLVP 1389 . : . : . : : : * : . * : : . : * : . : * : . : * : * : * :
gi 285814403 tpg DAA10297.1 gi 118841 sp P21951.1 DPOE_YEA gi 259485986 tpe CBF83469.1 gi 194214443 ref XP_001915746. Q07864-1	NNPKTSNPAGGQLFKITLPESVFLEEKENCTSIFNDENVLGVFE GTITPH 1434 NNPKTSNPAGGQLFKITLPESVFLEEKENCTSIFNDENVLGVFE GTITPH 1434 NG----HPS-VHLFKLTLSSEETFLREADKIHVLLQHPSVE GVYERNIPLN 1404 RS----NMVYNLYEYSPEDMYQEHEINEINTELSAAPDIE GVYE TQVPLL 1468 RS----NMVYNLYEYSPEDMYQEHEINEINAELSAAPDIE GVYE TQVPLL 1434 . : * : : : * : . : . : : . : . : * : * : * : :
gi 285814403 tpg DAA10297.1 gi 118841 sp P21951.1 DPOE_YEA gi 259485986 tpe CBF83469.1 gi 194214443 ref XP_001915746. Q07864-1	NVVLVDGVVDNLTVNTILTSALINDAEGSDLVNNNMGIKKDA----- 1770 NVVLVDGVVDNLTVNTILTSALINDAEGSDLVNNNMGIKKDA----- 1770 TVCVELEVRNLAINILTSSIIINEAEGADSLLAPS--DPSAE----- 1735 TVCVELDIQNLAVNTILQSHHVNDMEGADSMGVSDVIQQASLEDMITGN 1807 TVCVELDLQNLAVNTILQSHHVNDMEGADSMGIVSDVIQQASLEDMITGG 1773 . : * : : : * : * : * : * : * : * : * : * : * : * : .



Fig. 2K. CLUSTAL 2.0.12 multiple sequence alignment of epsilon polymerases (catalytic subunit)-Eukaryotes

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 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/ K¹ A/ _{T/S} V/I/ _M N/sy/FGLI/A Y⁹G	Alpha	L/ K¹ LTANSMY ⁸ GCLG
pol II	K/R/ _M / R¹ H/L/L/vXPXF/ Y⁷G	Beta	K/ _S / K¹ V/I/LD/fS/pE/ _K Y⁶I/L A/gT
pol III	N K_R¹ S/AHS/A/A/ Y⁷G/A/s	Gamma	A K ¹ I/VFNYG/ _A RI Y⁹GAG
pol IV	L/ _M /A K¹ L/ _M A/ _G XXXXK P⁹XG¹¹	Delta	L K ¹ I/VSANSV Y⁸GFTGA
pol V	A/s/ K¹ L/vAN/ _{M/Q} XXXXX	Epsilon	H K ¹ V/cILNSF Y⁸GYVMRKG
Viral	A/ _N / K¹ /RXF/IXXGXLY ⁹ 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 ATLANTNQLNRK ¹ ILINSLY ⁸ GALGNIH
Human HSV 1	800 AVLLDKQQAAIK ¹ VVCNSVY ⁸ GFTGVQH
<i>E. coli</i> DNA pol I	748 TVTSEQRRSAK ¹ AINFOGLIY ⁹ GMSAFLAR
<i>E. coli</i> DNA pol II	482 RQGNKPLSQALK ¹ IMNAFY ⁸ GVLGTTA
<i>E. coli</i> DNA pol III (alpha subunit)	663 YPDVQWQHESLK ¹ PVLEPTY ⁸ GIILYQE
<i>P. furiosus</i> DNA pol	477 KILLDYRQKAIK ¹ LLANSFY ⁸ GYYGYAK
Yeast Alpha DNA pol	933 RVQCDIRQQALK ¹ L TANSMY ⁸ GCLGYVN
Human Alpha DNA pol	939 ILQYDIRQKALK ¹ L TANSMY ⁸ GCLGSY
Human Gamma DNA pol	917 TTVGISREHAK ¹ IFNYGRIV ⁹ GAGQPFAER
Human Delta DNA pol (Catalytic subunit)	683 RQVLDGRQLALK ¹ VSANSVY ⁸ GFTGAQV
Human Epsilon DNA pol	798 EVLYDSLQLAHK ¹ CILNSFY ⁸ GYVMRKGAR
<i>A. thaliana</i> Delta DNA pol (Catalytic subunit)	679 KAVLDGRQLALK ¹ ISANSVY ⁸ GFTGATV

N.B: Some of the above polymerases did not align in T COFFEE advanced version. So the conserved regions were selected and aligned as suggested by Palanivelu [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 the maximum number of conserved motifs. ϵ polymerases showed long stretches of conserved motifs among them (Table 3).

A 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/IIXS/T	SP/TKQL	TGRLSS	PNLQNI	P DYS				
	QIEL/MRL/I/VL/MAH	IQGS/T	QVHDEL	VPL					
pol II	V/I/V/IG	V/I/LED	PGNHD	PQP	GHI/VH	V/NS			
pol III	GPGRGS	LL/I/FERFLNPER	PDX	DV/D F	LIKXFDFLGL	RPGP	IV/L/IYQEQQV/I		
pol IV	I/VIHV/ID	FFA	RGV/I/I	LSL/IDEA/G	GV/IG	VRLL/VGV/L/I			
pol V	D/EV/IWGV/IGR/K								
Viral	YNV/I	V/I/LN	ESI						
Table 3B. Eukaryotic polymerases									
Alpha	LFGK	LLHXF	PDK	GGLVLD/EP	ILLLDFNSLYPSIIQEQY/FNICFTTV/I	SRFY			
	VIYG	DTDSIMI	KKKYAA	KGLD	RRDW/Y/F				
Beta	CGSY/F	DXDVL	KIL	RRXDIR	GVLY				
Gamma	KLPH	GTITRRRAVEXTWLTASN	RLGSELK	F/LVGADVDSQ/EELWI					
	GT/SD	TKG	SRV/IN	VDY	LHL HDEV/IRYL	AFFSX DID	LRKE		
	TPSN								
Delta	LSFDIEC	FPE	DPVIQI	DPDI/V	GYN	FDL PYL	SYS/TLN		
	KEDVDLQN	YCLKD	EMARVTGXGP	RGG	GATVL/I	TLDF			
	SLYPSIMMAXNLCYXTL	GL/ILP	ARKR	GXLPC	V/VYG	DTDSVM			
	PIK/RLEFEKVYXPYLLIXKKRYAGL	KGI/L	LVTIK	V/LGDRVXY	EDP				
Epsilon	GWLTNM	VDYYFIQD/ED/EG	PYFXIA	KEDL	NHLL/V	DXREYDV	PYHIRLSID		
	DPVVL/MAFDIETTKXPLKF	PDXQDQIMMI							
	SYMIDGXGY/FLITNREIV	SEDIEDDFEF/YTPKPEY	I/VFNEF/WFEH						
	TYNGDFFDWPFV/I	EIGF	HMDCXWVXRDSYLPXGSXXXLKA	KLG YXP	XELDPE				
	YSVSDAVATYYLYMKYVHPFIFALCTIIPMEPDEXLRKGS/TGTLCE	XXLMVQAY/F							
	NII/L/VXPNK	HVLDSETVGGHVE	GVFRSDI/L	VTNF/Y					
	IYHLDVG/AXMYPNIXXNRQ	CAXCDFNKG	WXWRGEFXP	FXELS					
	V/ICQRENXFYVDTVR/KXFRDRRYD/EFKG	ARXLIEXIGRPLE	DTD GIWCXLPXS	FPE					
	L/ISYP	MLN	QYQYXTR/HSENS	TIFFEVDGPYXAMILPXS	KEE				
	KKRYAVFNE	DGS LAELKG	FEVKRRGELXLIKFQS/T	WLDVL	LIXENRSMSR/KXL				
	QKSTSIS/TTAK/RRLXEFLGDXMV	KDXGLXCR/KYIIS	P VTERA	IPLXIF	FLRKW				
	YYXERLGSXI	QKII	TPAALQXXXPVPRVXKHPDWL	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 (PXXG) 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 Mg^{2+} ion binds specifically to the beta and gamma phosphates of the dNTP as a bidentate.

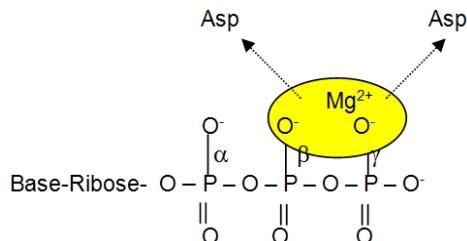


Fig. 3. dNTP-Mg ion complex where β - and γ -phosphates are shielded by the Mg ion
3.8 Functional domains of *E. coli* DNA polymerase I

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 $3' \rightarrow 5'$ exonuclease activity (proof-reading function) and the $5' \rightarrow 3'$ 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 $5' \rightarrow 3'$ exonuclease activity. Further analysis of the enzyme has shown that the amino acids from 1-324 perform the $5' \rightarrow 3'$ exonuclease activity (DNA repair function); 324-517 perform the $3' \rightarrow 5'$ 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).

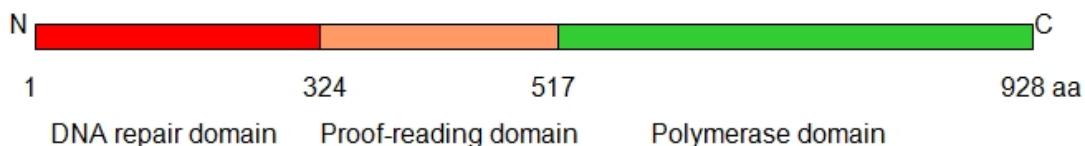


Fig. 4. Dissection of the 3 functional domains of *E. coli* DNA polymerase I

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 KY/I/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].

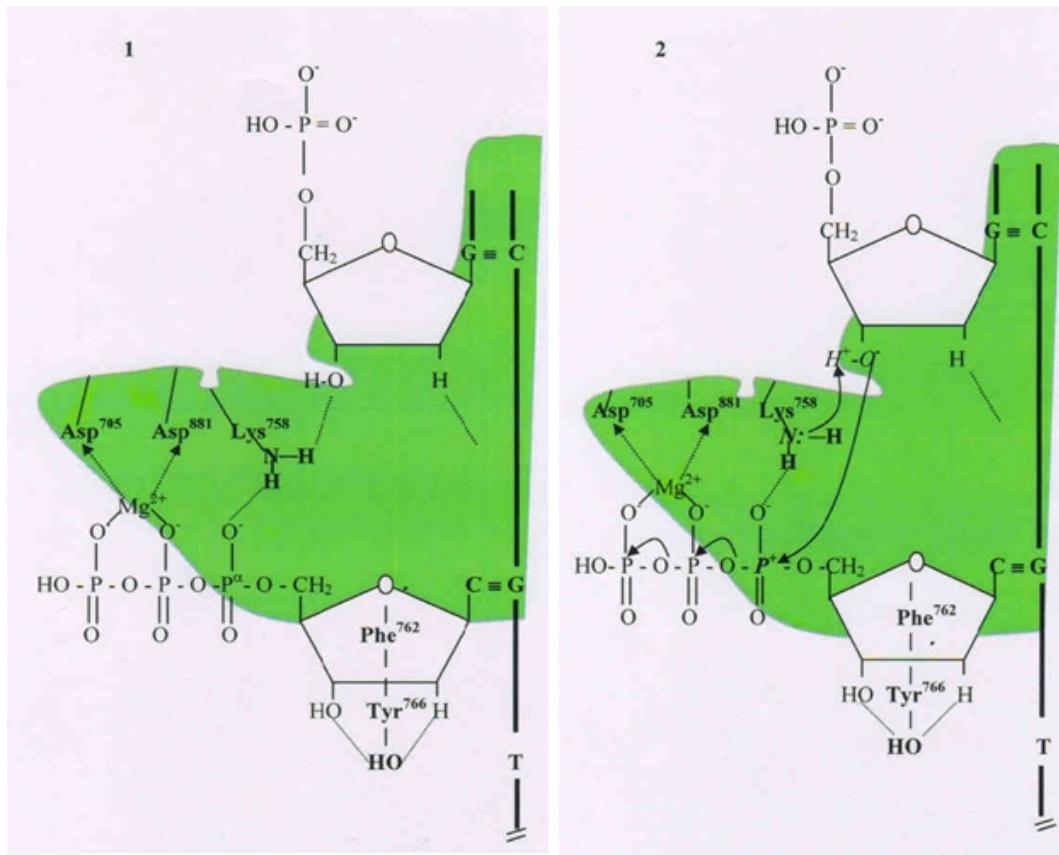
In contrast, Tyr is found in the active site of topoisomerases, which actually involve in breaking of phosphodiester bonds. Interestingly, Tyr is conserved in all topoisomerases. In fact, the active site Tyr sometimes makes a covalent bond with the 5' end of the DNA. It is interesting to note that these 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 site directed mutagenesis (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

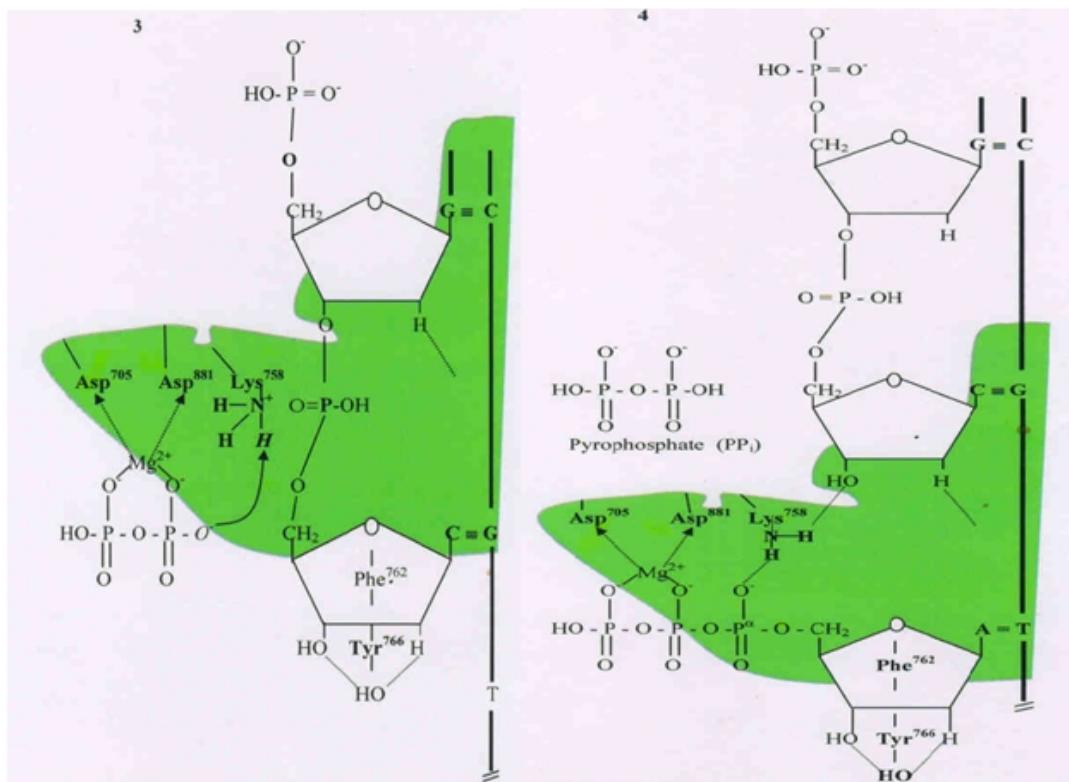
analysis by SDM by Doublie and Ellenberger [18] and Astatke et al. [19] have shown that the critical Tyr may possibly be involved in template recognition and dNTP selection in DNA polymerases. It is also known that a highly conserved Tyr residue in reverse transcriptase controls substrate selection. 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; an 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 highly conserved and found to be a common pair in different types of DNA polymerases (Table 2).

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

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

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





Figs. 5.1 - 5.4. Proposed mechanism for polymerization reaction of *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, subsite 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).

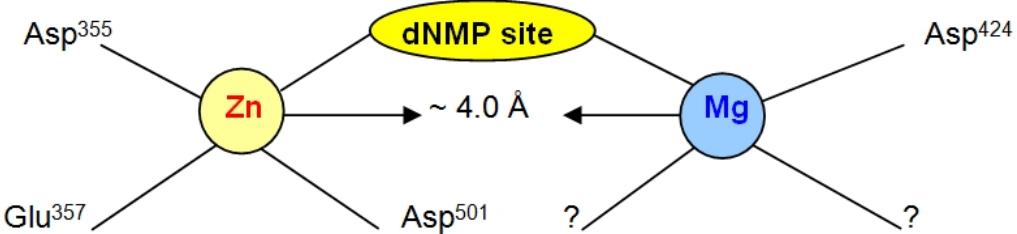


Fig. 6. Schematic diagram showing the subsites A and B of proof-reading activity of *E. coli* DNA polymerase I

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

- a) 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.
- b) 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.

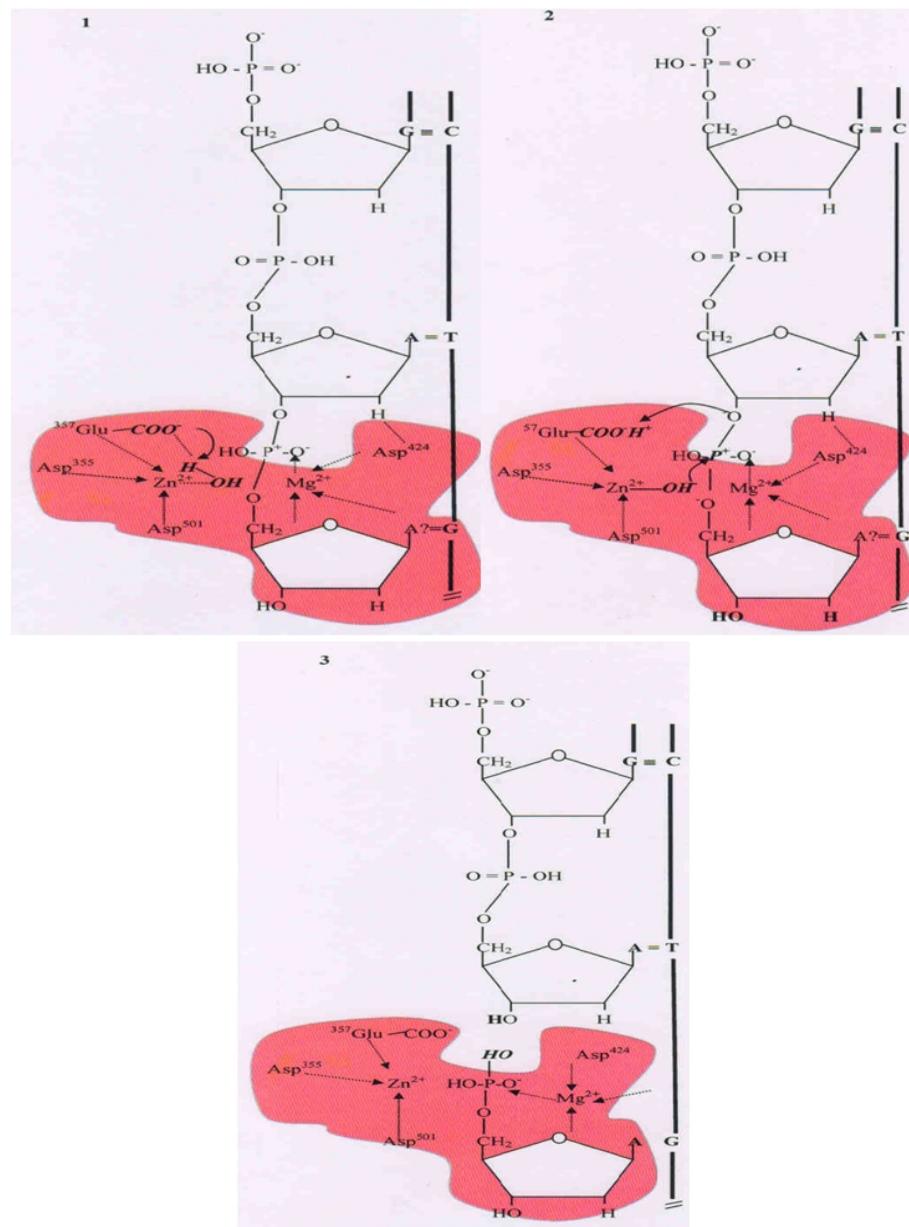
Generally, in all Zn containing enzymes, a water molecule is invariably coordinated to the Zn atom as the fourth ligand, the other three ligands could be contributed by carboxylic acids (Glu, Asp), His or Cys.

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 proof-reading function of *E. coli* DNA polymerase I

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.

4. 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. DNA Replication. *J Biol Chem.* 1988;263:1-4.
2. Steitz TA. DNA Polymerases: Structural Diversity and Common Mechanisms. *J. biol chem.* 274:1999;17395-17398.
3. Ohmori H, Friedberg EC, Fuchs RP, Goodman MF, Hanaoka F, et al. The Y-family of DNA polymerases. 2001; *Mol Cell*, 8:7-8.
4. Palanivelu P. Multiple sequence analysis of polygalacturonases and invertases and phase shift in conserved motifs. *Indian J Biotechnol.* 2007;6:24-30.
5. Joyce CM, Steitz TA. DNA polymerase I: from crystal structure to function via genetics. *Trends Biochem Sci.* 1987;12:288-292.
6. Beese LS, Steitz TA. Structural basis for the 3'- 5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* 1991;10:25-33.
7. Rush J, Konisberg WH. Photoaffinity labeling of the Klenow fragment with 8-azido-dATP. *J Biol Chem.* 1990;265:4821-4827.
8. Basu A, Modak MJ. Affinity labeling of *E. coli* DNA polymerase I by pyridoxal 5'-phosphate. *Biochemistry.* 1987;26:1704-1709.
9. Monaghan N, Hay RT. Pyridoxal 5'-phosphate inhibition of adenovirus DNA polymerase. *J Biol Chem.* 1996;271:24242-24248.
10. Zaldivar MJ, Bull P, Venegas A, Valenzuela P. Inactivation of rat liver RNA polymerases I and II and yeast RNA polymerase I by pyridoxal 5'-phosphate. Evidence for the participation of lysyl residues at the active site. *Biochemistry.* 1975;14:4907-4911.
11. Palanivelu P. Enzymes, Ribozymes and DNAzymes. Twentyfirst Century Publications, Palkalai Nagar, Madurai, India; 2007.
12. Blasco MA, Lázaro JM, Blanco L , Salas M. Phi 29 DNA polymerase active site. The conserved amino acid motif "Kx3NSxYG" is involved in template-primer binding and dNTP selection. *Proc Natl Acad Sci (USA).* 1997;94:1619-1622.

13. Sriskanda V, Schwer B, Ho CK, Shuman S. 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. 1999;27:3953-3963.
14. Tomkinson AE, Totty NF, Ginsburg M, Lindahl T. Location of the active site for enzyme-adenylate formation in DNA ligases. Proc Natl Acad Sci (USA). 1991;88:400-404.
15. Redinbo MR, Stewart L, Kuhn P, Champoux JJ, Hol WGJ. Crystal Structures of Human Topoisomerase I in Covalent and Noncovalent Complexes with DNA. Science. 1998;279:1504-1513.
16. Kaushik N, Pandey VN, Modak MJ. Significance of the O-helix residues of *Escherichia coli* DNA polymerase I in DNA synthesis; dynamics of the dNTP binding pocket. Biochemistry. 1996;35:7256-66.
17. Singh K, Modak MJ. A unified DNA- and dNTP binding mode for DNA polymerases. Trends Biochem Sci. 1998;23:277-280.
18. Doublie S, Ellenberger T. The mechanism of action of T7 DNA polymerase. Curr Opin Struct Biol. 1998;8:704-712.
19. Astatke M, Grindley ND, Joyce CM. How *E. coli* DNA polymerase I (Klenow fragment) distinguishes between deoxy- and dideoxynucleotides. J Mol Biol. 1998;278:147-165.
20. Ponamarev MV, Longley MJ, Nguyen D, Kunkel TA, Copeland WC. Active site mutation in DNA polymerase γ associated with progressive external ophthalmoplegia causes error-prone DNA synthesis. J Biol Chem. 2002;277:15225-15228.
21. Doublie S, Tabor S, Long AM, Richardson CC, Ellengerger T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. Nature. 1998;391:251-258.
22. Joyce CM. Choosing the right sugar: How polymerases select a nucleotide substrate. Proc Natl Acad Sci (USA). 1997;94:1619-1622.
23. Derbyshire V, Freemont PS, Sanderson MR, Beese L, Friedman JM, Joyce CM, Steitz TA. Genetic and crystallographic studies of the 3,5'-exonucleolytic site of DNA polymerase I. Science. 1988;240:199-201.
24. Derbyshire V, Grindley ND, Joyce CM. 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. 1991;10:17-24.
25. Vallee BL, Auld DS. Active-site zinc ligands and activated H₂O of zinc enzymes. Proc Natl Acad Sci (USA). 1990;87:220-224.
26. Vallee BL, Auld DS. Active zinc binding sites of zinc metalloenzymes 1: Matrix Suppl. 1992;5:19.
27. Mizrahi V, Benkovic P, Benkovic SJ. Mechanism of DNA polymerase I: exonuclease/polymerase activity switch and DNA sequence dependence of pyrophosphorolysis and misincorporation reactions. Proc Natl Acad Sci, (USA). 1986;83:5769-5773.

© 2013 Palanivelu; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here:

<http://www.sciedomain.org/review-history.php?iid=231&id=3&aid=1432>