The EMBO Journal vol.6 no.13 pp.4219–4225, 1987

Structural and functional relationships between prokaryotic and eukaryotic DNA polymerases

Antonio Bernad, Angel Zaballos, Margarita Salas and Luis Blanco

Centro de Biología Molecular (CSIC–UAM), Universidad Autónoma, Canto Blanco, 28049 Madrid, Spain

Communicated by M.Salas

The *Bacillus subtilis* phage φ29 DNA polymerase, involved in protein-primed viral DNA replication, was inhibited by phosphonoacetic acid (PAA), a known inhibitor of α-like DNA polymerases, by decreasing the rate of elongation. Three highly conserved regions of amino acid homology, found in several viral α-like DNA polymerases and in the φ29 DNA polymerase, one of them proposed to be the PAA binding site, were also found in the T4 DNA polymerase. This prokaryotic enzyme was highly sensitive to the drugs aphidicolin and the nucleotide analogues butylanilino dATP (BuAdATP) and butylphenyl dGTP (BuPdGTP), known to be specific inhibitors of eukaryotic α-like DNA polymerases. Two potential DNA polymerases from the linear plasmid pGKL1 from yeast and the S1 mitochondrial DNA from maize have been identified, based on the fact that they contain the three conserved regions of amino acid homology. Comparison of DNA polymerases from prokaryotic and eukaryotic origin showed extensive amino acid homology in addition to highly conserved domains. These findings reflect evolutionary relationships between hypothetically unrelated DNA polymerases.

Key words: α-like DNA polymerases/amino acid homology/bacteriophage φ29/bacteriophage T4/linear plasmids

Introduction

Prokaryotic and eukaryotic DNA polymerases of viral and cellular origin differ in many properties such as their size, ability to interact with accessory proteins and biological role (Kornberg, 1980). However, the template-directed copying of DNA is identical in all cases, and it is therefore likely that functional and even structural similarities can be found in apparently unrelated DNA polymerases.

The linear DNA of the *Bacillus subtilis* phage φ29 replicates by a protein-priming mechanism in which a unique virally coded DNA polymerase, protein p2, catalyses both the formation of the initiation complex between the terminal protein p3 and dAMP and its further elongation to produce unit-length φ29 DNA (Blanco and Salas, 1985a). In addition, the φ29 DNA polymerase has a 3′→5′ exonuclease activity on single-stranded DNA (Blanco and Salas, 1985b). Adenovirus DNA replicates by a similar protein-priming mechanism which also requires a virally coded DNA polymerase (Nagata *et al.*, 1983).

The φ29 DNA polymerase is inhibited by aphidicolin and the nucleotide analogues butylanilino dATP (BuAdATP) and butylphenyl dGTP (BuPdGTP) (Blanco and Salas, 1986) which, in addition to phosphonoacetic acid (PAA), are known inhibitors of the eukaryotic DNA polymerase α (reviewed by Huberman, 1981; Khan *et al.*, 1984, 1985) and other α-like DNA polymerases of viral origin such as those of herpes simplex (HSV) and human cytomegalovirus (HCMV) (Knopf, 1979; Nishiyama *et al.*, 1983), African swine fever (ASFV) (Moreno *et al.*, 1978), vaccinia (Moss and Cooper, 1982; Sridhar and Condit, 1983) and Epstein–Barr (EBV) (Kallin *et al.*, 1985). Three conserved amino acid domains have been recently reported to be present in the C-terminal portion of several viral α-like DNA polymerases and in the φ29 DNA polymerase (Larder *et al.*, 1987; Knopf, 1987). Mutations conferring altered sensitivity to antiviral pyrophosphate (PP) analogues or aphidicolin were precisely located in or near these conserved regions (Gibbs *et al.*, 1985; Earl *et al.*, 1986; Larder *et al.*, 1987; Knopf, 1987) which were proposed to form part of the catalytic domains of the enzyme (Gibbs *et al.*, 1985).

In this paper we report the inhibition of φ29 DNA–protein p3 replication by PAA. The drug inhibited the rate of elongation but not the initiation step in replication. We have also found that the prokaryotic phage T4 DNA polymerase is strongly inhibited by nucleotide analogues and aphidicolin and, in agreement with this finding, it contains the three conserved amino acid domains recently described. In addition two potential DNA polymerases of the terminal protein-containing linear plasmid pGKL1 from yeast (Gunge *et al.*, 1981) and S1 mitochondrial DNA from maize (Pring *et al.*, 1977) were identified based on the presence of the DNA polymerase conserved domains. Amino acid comparisons between several prokaryotic and eukaryotic DNA polymerases showed that, in addition to highly conserved domains, there is extensive amino acid sequence conservation, and therefore it is likely that this family of polymerases evolved from a common ancestor.

Results and discussion

Inhibition of φ29 DNA–protein p3 replication in vitro by PAA

Figure 1A shows that PAA strongly inhibited φ29 DNA–protein p3 replication *in vitro* using purified DNA polymerase p2 and terminal protein p3, the inhibition being 50% at a drug concentration of 70 μM. Figure 1A also shows that a concentration of PP 10-fold higher than in the case of the PAA was needed to produce 50% inhibition. Taking into account that protein p2 is involved both in initiation and elongation events, the effect of PAA on each of these two steps was studied. No effect of PAA (Figure 1B) or PP (not shown) on the amount of p3–dAMP initiation complex formed was observed. Taking into account these results, as well as the fact that PAA did not inhibit the transition from the p3–dAMP initiation complex to the first elongation product, p3–$(dAMP)_2$, determined as described by Blanco *et al.* (1986), nor the binding of protein p2 to either double- or single-stranded DNA (not shown), determined by a gel retardation assay as described by Carthew *et al.* (1985), the effect of the drug on the rate of elongation was determined. Figure 2A shows that, when the DNA synthesized *in vitro* in the presence of increasing amounts of PAA was run on an alkaline agarose gel, a decrease in the length of the DNA was observed. PP also decreased the length of the DNA synthesized, although as ex-

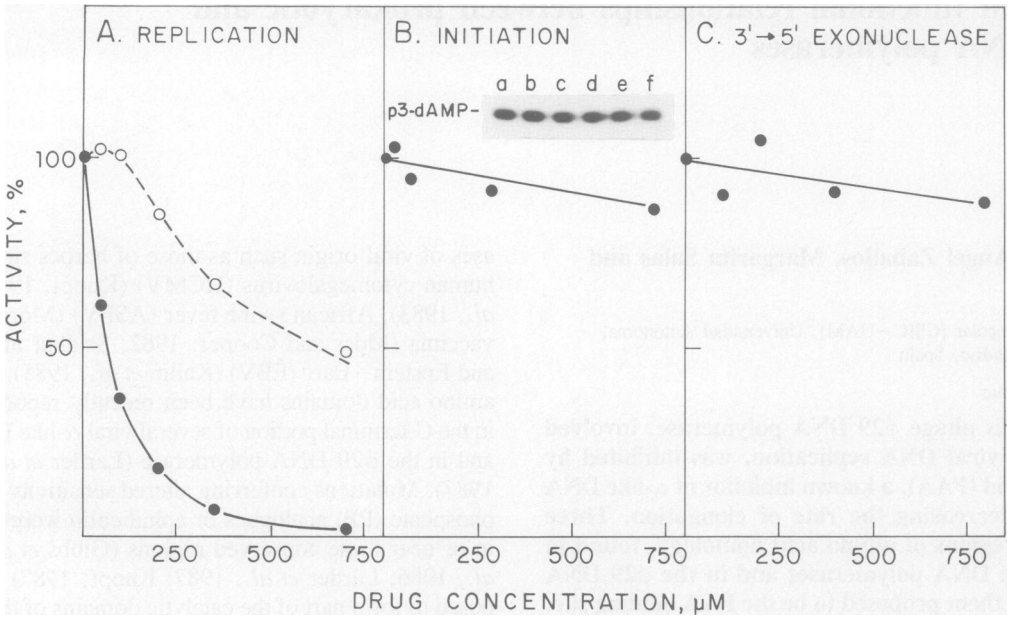


Fig. 1. Effect of PAA on ϕ29 DNA−protein p3 replication and on the 3′→5′ exonuclease activity of protein p2. **A**. The replication assay was carried out as described in Materials and methods in the absence or presence of the indicated amounts of PAA (●—●) or PP (○—○). After incubation for 4 min at 30°C the samples were processed as described in Materials and methods. 100% ϕ29 DNA−p3 replication corresponds to 39 pmol of dAMP incorporated. **B**. As in A, except that, after 7.5 min at 30°c, the samples were treated with micrococcal nuclease to degrade the elongation product and analyze the p3−dAMP initiation complex formed. The amount of PAA used was: **a**, none; **b**, 21 μM; **c**, 71 μM; **d**, 142 μM; **e**, 284 μM; **f**, 714 μM. 100% of p3−dAMP complex formation corresponds to 160 fmol. **C**. Protein p2 was incubated with heat-denatured [3′-^{32}P]DNA in the absence or presence of the indicated amounts of PAA as described in Materials and methods, and the ethanol-soluble material was counted by Cerenkov radiation. 100% activity corresponds to 1400 c.p.m. of [^{32}P]DNA hydrolyzed.

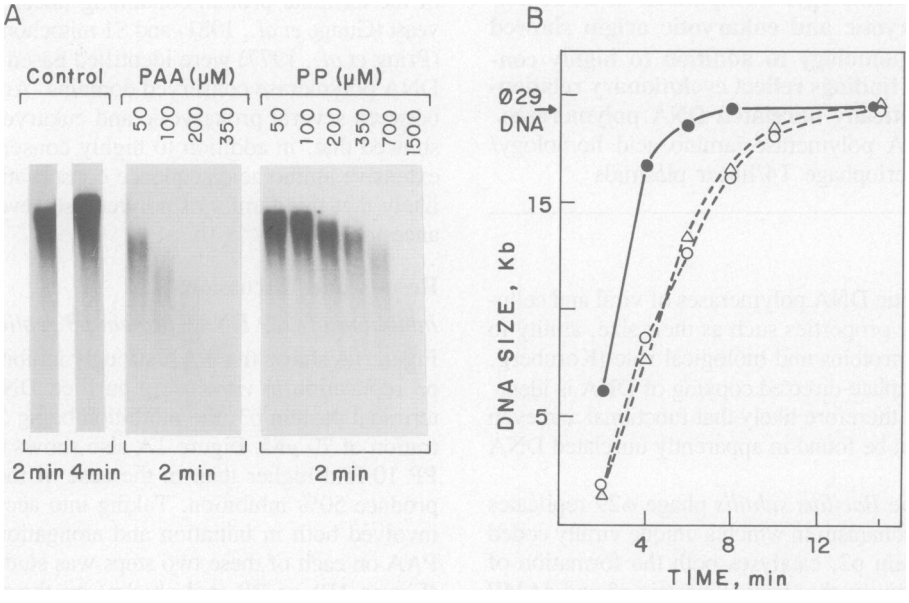


Fig. 2. Effect of PAA on the rate of elongation of ϕ29 DNA−protein p3. **A**. The replication assay was carried out as described in Materials and methods in the absence or presence of the indicated amounts of PAA. After incubation for the indicated times at 30°C the samples were subjected to alkaline agarose gel electrophoresis as described in Materials and methods. **B**. The size of the DNA synthesized as determined by alkaline agarose gel electrophoresis in the absence (●—●) or presence of 100 μM PAA (○--○) or 1000 μM PP (△--△) is represented as a function of the incubation time.

pected from the results shown in Figure 1A, a concentration about 10-fold higher than in the case of PAA was needed to produce a similar effect (Figure 2A). As shown in Figure 2B, addition of 0.1 mM PAA or 1 mM PP produced a 2-fold decrease in the rate of elongation.

Neither PAA (Figure 1C) nor PP (not shown) inhibited to a high extent the activity of the 3′→5′ exonuclease on single-stranded DNA of the ϕ29 DNA polymerase. A concentration of 1.5 mM PAA was needed to produce 40% inhibition (not shown). This is in contrast to the polymerase and 3′→5′ exonuclease activities of the HSV type I DNA polymerase which are both inhibited to a similar extent by PAA (Knopf, 1987).

Table I. Location of the consensus motifs in several DNA polymerases

DNA polymerase	Size (aa)	Location (%)[a]		
		SLYP	NS-YG-F	Y-DTDS
S1 mit[b]	917	53.4	68.0	78.2
pGKL1[b]	995	65.7	79.5	87.9
Adeno	1056	51.8	66.5	82.4
ϕ29	575	44.1	67.8	79.4
PRD1	553	40.6	62.7	77.5
T4	896	46.0	63.0	69.0
Vaccinia	938	56.3	68.5	77.5
HSV 2	1240	58.6	66.4	72.0
VZV	1194	57.5	65.5	71.3
EBV	1015	58.0	67.7	74.4
HCMV	1242	58.0	65.9	73.3

[a]Location is expressed as % from the N termini.
[b]Indicates predicted DNA polymerases.

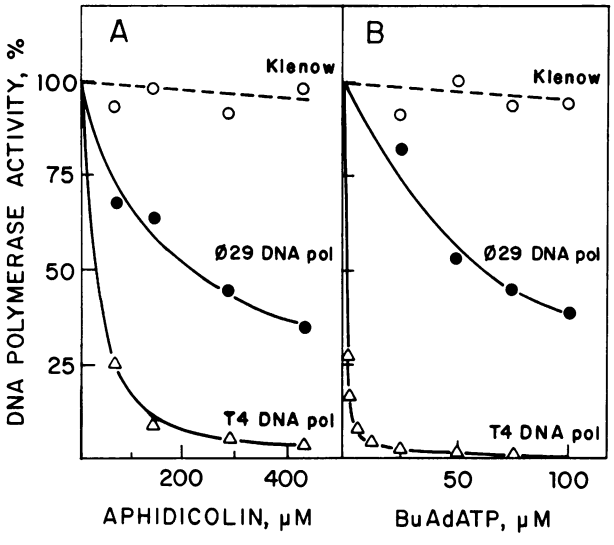


Fig. 3. Effect of aphidicolin and BuAdATP on the replication of activated DNA by the ϕ29 and T4 DNA polymerases and the Klenow enzyme. The replication assay was as described in Materials and methods using activated DNA as template and the ϕ29 DNA polymerase, the phage T4 DNA polymerase or the Klenow fragment of *E.coli* DNA polymerase I in the absence or presence of the indicated amounts of aphidicolin (**A**) or BuAdATP (**B**). After incubation for 4 min at 30°C the reaction was stopped and the samples were filtered through Sephadex G-50 spun columns as described in Materials and methods.

Search for related structural domains in other prokaryotic DNA polymerases

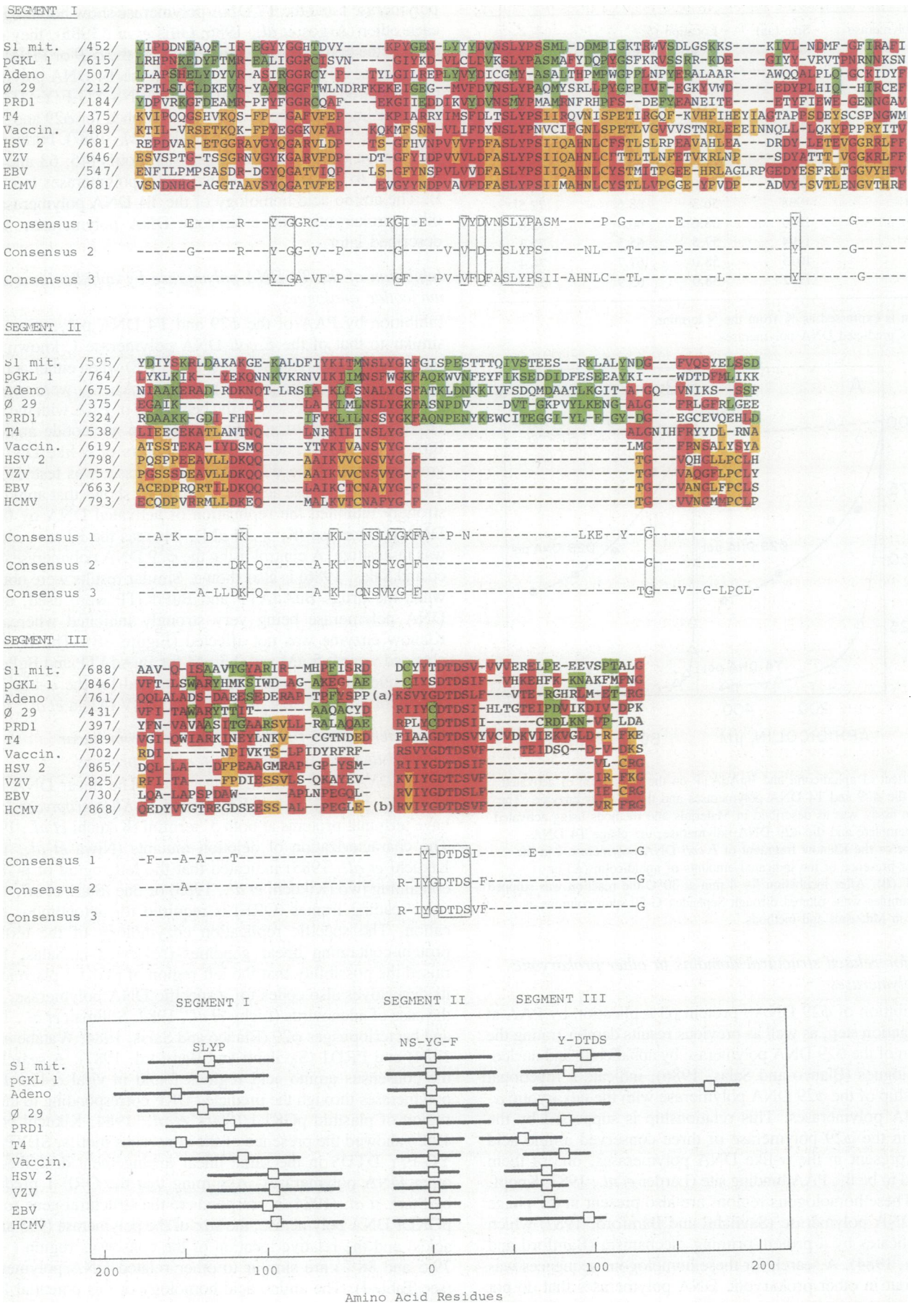
The inhibition of ϕ29 DNA−protein p3 replication by PAA at the elongation step, as well as previous results demonstrating the inhibition of the ϕ29 DNA polymerase by aphidicolin and nucleotide analogues (Blanco and Salas, 1986), indicate a functional relationship of the ϕ29 DNA polymerase with the eukaryotic α-like DNA polymerases. This relationship is supported by the finding in the ϕ29 polymerase of three conserved amino acid regions present in the α-like DNA polymerases, one of them proposed to be the PAA binding site (Larder *et al.*, 1987; Knopf, 1987). These homologous regions are also present in the phage PRD1 DNA polymerase (Savilahti and Bamford, 1987) which also replicates by a protein-priming mechanism (Bamford and Mindich, 1984). A search for these homologous sequences was carried out in other prokaryotic DNA polymerases that do not use a protein as primer. Whereas the *Escherichia coli* DNA polymerase I and the T7 DNA polymerase show homology with each other (Argos *et al.*, 1986; Ollis *et al.*, 1985), they do not contain the homologous regions found in the ϕ29 or PRD1 DNA polymerases. However, we found that the T4 DNA polymerase has the three consensus motifs: SLYP/NS-YG-F/Y-DTDS, in the same linear arrangement as was described for ϕ29 and several viral α-like DNA polymerases (Larder *et al.*, 1987). Furthermore, the relative location of the consensus motifs (46, 63 and 69%) is similar to that of other related DNA polymerases (see Table I). The amino acid homology of the T4 DNA polymerase with other prokaryotic and eukaryotic DNA polymerases will be described later.

Inhibition of the T4 DNA polymerase by aphidicolin and nucleotide analogues

Inhibition by PAA of the ϕ29 and T4 DNA polymerases was similar to that of the *E.coli* DNA polymerase I, known to be rather insensitive (Helgstrand *et al.*, 1978) when either activated DNA or the template-primer poly dT-$(dA)_{12-18}$ were used (results not shown). In view of the behaviour of PAA with the above DNAs, the effect of aphidicolin and the nucleotide analogues BuAdATP and BuPdGTP, also inhibitors of α-like DNA polymerases and of the ϕ29 DNA polymerase, was tested on the T4 DNA polymerase activity. Figure 3A shows that aphidicolin strongly inhibited the replication of activated DNA by the T4 DNA polymerase (K_i = 30 μM), whereas the Klenow enzyme was insensitive. A control using the ϕ29 DNA polymerase (Blanco and Salas, 1986) is also shown. Similar results were obtained when the drugs BuAdATP and BuPdGTP were used, the T4 DNA polymerase being very strongly inhibited whereas the Klenow enzyme was not affected (Figure 3B). The K_i values obtained were 0.5 μM and 0.3 μM for BuAdATP and BuPdGTP respectively, very close to those obtained for the α-like DNA polymerases from EBV, vaccinia and HSV (Khan *et al.*, 1984).

Characterization of DNA polymerases from linear plasmid-like DNAs containing terminal proteins

Linear DNA killer plasmids from yeast. The linear DNA killer plasmids, pGKL1 and pGKL2 found in *Kluyveromyces lactis*, have terminal proteins at both 5′ termini (Kikuchi *et al.*, 1984). The characterization of deletion mutants (Niwa *et al.*, 1981; Kikuchi *et al.*, 1985) indicated that the left region of pGKL1, containing two (Kikuchi *et al.*, 1985) or one (Stark *et al.*, 1984) open reading frames (ORF) is required for autonomous replication. The specific replication enzymology of the terminal protein-containing linear genomes (reviewed in Salas, 1983) raised the possibility that the left region of pGKL1 plasmid and its derivatives also codes for a specific DNA polymerase, as in the case of adenovirus (Lichy *et al.*, 1982; Stillman *et al.*, 1982) and bacteriophages ϕ29 (Blanco and Salas, 1984; Watabe *et al.*, 1984) and PRD1 (Savilahti and Bamford, 1987). A search for the consensus amino acid regions found in viral α-like DNA polymerases through the predicted ORF corresponding to the left region of plasmid pGKL1 (Stark *et al.*, 1984; Kikuchi *et al.*, 1985) showed the presence of the consensus motifs, SLYP/NS-YG-F/Y-DTDS in the same linear arrangement described for other DNA polymerases. Assuming that the ORF 1 predicted by Stark *et al.* (1984) corresponds to the structural gene for the pGKL1 DNA polymerase, the size of the polymerase (995 amino acids) and the relative location of the consensus region (~66, 79.5 and 88%) are similar to other related DNA polymerases (see Table I). The amino acid homology of this potential DNA


SEGMENT I
S1 mit. /452/
pGKL 1 /615/
Adeno /507/
Ø 29 /212/
PRD1 /184/
T4 /375/
Vaccin. /489/
HSV 2 /681/
VZV /646/
EBV /547/
HCMV /681/
Consensus 1
Consensus 2
Consensus 3
SEGMENT II
S1 mit. /595/
pGKL 1 /764/
Adeno /675/
Ø 29 /375/
PRD1 /324/
T4 /538/
Vaccin. /619/
HSV 2 /798/
VZV /757/
EBV /663/
HCMV /793/
Consensus 1
Consensus 2
Consensus 3
SEGMENT III
S1 mit. /688/
pGKL 1 /846/
Adeno /761/
Ø 29 /431/
PRD1 /397/
T4 /589/
Vaccin. /702/
HSV 2 /865/
VZV /825/
EBV /730/
HCMV /868/
Consensus 1
Consensus 2
Consensus 3
SEGMENT I
SEGMENT II
SEGMENT III
SLYP
NS-YG-F
Y-DTDS
S1 mit.
pGKL 1
Adeno
Ø 29
PRD1
T4
Vaccin.
HSV 2
VZV
EBV
HCMV
200
100
0
100
200
Amino Acid Residues

polymerase with other viral DNA polymerases will be described below. The fact that the linear plasmid pGKL2 does not require the presence of pGKL1 for is maintenance suggests that pGKL2 also codes for its own DNA polymerase. If so, it would be interesting to compare these two DNA polymerases when the sequence of plasmid pGKL2 becomes available.

S1 mitochondrial DNA from maize. Mitochondria from S-male-sterile cytoplasm of maize contain DNA−protein complexes designated S1 and S2 (Pring *et al.*, 1977). These complexes consist of double-stranded linear DNA with proteins covalently attached to the 5′ termini (Kemble and Thompson, 1982). Based on their structural organization and their viral-like characteristics, Paillard *et al.* (1985) proposed that S1 and S2 code for functions involved in their maintenance and replication. A computer search for the consensus amino acid regions found in viral α-like DNA polymerases through the potential ORF of S1 and S2 DNA (Paillard *et al.*, 1985; Levins and Pring, 1979) indicated that the ORF 3 from S1 DNA contains the three consensus regions SLYP/NS-YG-F/Y-DTDS described before; as shown in Table I, these regions are located about 53, 68 and 78% from the amino terminus, in good agreement with their location in other related viral DNA polymerases. These results strongly suggest that S1 DNA encodes its own DNA polymerase. The amino acid homology of this new DNA polymerase with other viral DNA polymerases will be described below.

In addition to bacteriophages φ29, PRD1 and related phages, adenovirus and plasmids pGKL1, pGKL2 and pSKL (Kitada and Hishinuma, 1987) from yeast and S1 and S2 from *Zea mays*, other protein-containing linear DNA molecules have been found in several organisms including Cp-1 and related phages of *Streptococcus pneumoniae* (García *et al.*, 1983; Escarmís *et al.*, 1985), plasmids pSLA1 and pSLA2 of *Streptomyces rochei* (Hirochika and Sakaguchi, 1982), mitochondrial plasmid from *Brassica* (Palmer *et al.*, 1983) and the linear DNA from the fungus *Ascobolus immersus* (Francou, 1981). In some cases the terminal protein was shown to be required for the initiation of replication. It is likely that a specific DNA polymerase, containing the consensus amino acid regions found in many viral DNA polymerases, is also required in each case.

Amino acid homologies between DNA polymerases reflect evolutionary relationships

The amino acid sequences of the different DNA polymerases indicated in Materials and methods were aligned to generate maximum cross-homology. The three segments shown in Figure 4 represent the most conserved regions of the polymerases; outside these segments, the sequences show less homology, and a higher number of insertions and deletions are necessary for alignment. The addition of the S1 mitochondrial DNA, pGKL1, T4, *Varicela zoster* virus (VZV) and HCMV DNA polymerase sequences to the list of homologous polymerases described previously (Larder *et al.*, 1987; Kouzarides *et al.*, 1987; Savilahti and Bamford, 1987), defines more precisely the boundaries of the consensus regions and the extent of conservation, and makes it possible to distinguish consensus sequences accounting for replication strategies (see legend to Figure 4 for details). Figure 4 also shows that these three segments of amino acid homology are located in the same linear arrangement in each protein. As already indicated, it is interesting to note that the main motif of the consensus sequences of each segment is located at a similar relative position, taking into account the size of the different DNA polymerases considered (Table I).

Although we can define segments of conserved sequences, considerable variation remains among the DNA polymerases considered in Figure 4. The amino-terminal half of the different DNA polymerases described here shows less cross-homology than the C-terminal region. These non-homologous regions could either represent alternative primary sequences capable of attaining similar three-dimensional structures, or they could have specific functions in the DNA polymerases, such as interaction with other replication proteins. It is possible that, in the case of genomes that replicate by a protein-priming mechanism, the insertion present in the neighbourhood of a hypothetically catalytic domain in segment II (see Figure 4) might be designed for proper interaction with the terminal protein allowing its deoxynucleotidilylation.

An overall view of the consensus sequences shown in Figure 4 indicates that, although the amino acid homology in each segment is more extensive in the group of Herpes virus (HSV, VZV, EBV and HCMV), in keeping with their phylogenetic distances, and between the DNA polymerases of terminal protein containing genomes (consensus 1), probably due to mechanistic similarities in their replication strategies, there is, in addition to the highly conserved regions (consensus 2), extensive amino acid homology between the different DNA polymerases.

Due to the fact that the length of the polypeptides of segments II and III differs considerably in each DNA polymerase, and therefore the degree of homology cannot be easily quantified, we considered the homologies in segment I as an index of evolutionary relationships between these DNA polymerases. Table II shows the percentage of identical residues in segment I between each pair of DNA polymerases, the maximal value being indicated in each case. The values obtained show that the DNA polymerases of the Herpes virus group, with an average value of about 45%, are very proximal in evolution; among them, the DNA polymerases of HSV 2 and VZV (60% homology) would be the most closely related. Unexpectedly the DNA polymerases most related to the prokaryotic T4 DNA polymerase are eukaryotic viral α-like polymerases, in particular that of vaccinia virus (32% homology). Furthermore the percentage of homology between the vaccinia DNA polymerase and those of the Herpes virus group is also similar in the case of the T4 DNA polymerase and, in both cases, higher than the values obtained when compared with the other DNA polymerases. The detailed comparison of segment I of vaccinia and T4 DNA polymerases (see Figure 4) strongly suggests that they derived from a common ancestor.

Fig. 4. Homology segments among several DNA polymerases of prokaryotic and eukaryotic origin. Alignment was carried out as described in Materials and methods. The residues that were identical in at least two polypeptides were boxed in colour. Green boxes indicate residues that are homologous only between DNA polymerases of the terminal protein containing genomes (S1 mit, pGKL1, Adeno, φ29 and PRD1). Yellow boxes indicate residues that are homologous only between the rest of DNA polymerases (T4, vaccinia, HSV 2, VZV, EBV and HCMV). Red boxes indicate residues that are homologous between these two types of DNA polymerases. Numbers at the left indicate the first residue in each sequence. A residue was considered consensus if it was present in more than half of the polypeptides considered. Consensus 1 indicates the consensus residues between DNA polymerases of terminal protein containing genomes. Consensus 3 indicates the consensus residues between the remaining DNA polymerases. Consensus 2 indicates the consensus residues between all DNA polymerases. The consensus amino acids which are identical in the three kinds of consensus considered are boxed. (**a**) and (**b**) in segment III indicate insertions of 78 and 13 amino acid residues respectively. The spatial relationship between segments I−III (black bars) of each DNA polymerase are shown at the bottom. The tyrosine residue in the consensus motif NS-YG-F of segment II was taken as residue 0 in each case. Boxes indicate the main consensus motif in each segment.

Table II. Amino acid homology between prokaryotic and eukaryotic DNA polymerases

	S1 mit	pGKL1	Adeno	φ29	PRD1	T4	Vaccinia	HSV 2	VZV	EBV	HCMV
S1 mit	100	25.6	17.9	**25.6**	20.5	14.1	16.7	20.5	16.7	15.4	14.1
pGKL1	26.2	100	22.5	21.2	**26.2**	12.5	16.2	15.0	16.2	18.7	17.5
Adeno	17.1	21.9	100	**23.2**	18.3	8.5	7.3	17.1	17.1	13.4	17.1
φ29	24.1	20.5	22.9	100	**27.7**	15.7	18.1	20.5	19.3	18.1	20.5
PRD1	22.5	26.2	20.0	**28.7**	100	15.0	15.0	23.7	20.0	18.7	15.0
T4	11.9	11.9	8.3	15.5	14.3	100	**32.1**	17.9	20.2	19.0	22.6
Vaccinia	15.1	15.1	7.0	17.4	13.9	**31.4**	100	26.7	20.9	23.3	19.8
HSV 2	19.0	14.3	17.9	20.2	23.8	19.0	28.6	100	**59.5**	46.4	42.3
VZV	15.7	16.9	18.1	19.3	19.3	21.7	21.7	**60.2**	100	48.2	44.6
EBV	13.8	14.9	12.6	17.2	14.9	19.5	23.0	44.8	44.8	100	**46.0**
HCMV	13.2	16.9	16.9	21.7	15.7	22.9	20.5	43.4	44.6	**48.2**	100

Numbers indicate the percent relationship homologies in segment I (Figure 4) between each pair of DNA polymerases; reciprocal values are also indicated due to the different length of the sequence considered as 100% in each comparison. Figures printed in bold represent the maximal value obtained for each DNA polymerase.

The predicted DNA polymerases of the linear plasmids pGKL1 and S1 mitochondrial DNA, both containing terminal proteins, have the maximal homology between them, and with the DNA polymerases of the bacteriophages PRD1 and φ29 respectively, in agreement with a possible protein-priming mechanism of replication. The results presented in this paper provide genetic evidence suggesting a phylogenetic relationship between viruses and plasmid-like DNAs. The homology between the DNA polymerase of adenovirus and that of phages φ29 and PRD1, that are known to replicate by a protein-priming mechanism, is about 20–23%, not much higher than the homology with other DNA polymerases which do not use proteins as primers. The low values indicate that they are not very closely related, probably due to a strong divergence based on different host specificity. However, as shown in Figure 4, it is possible to distinguish particular consensus sequences (consensus 1 versus 3) in the three segments considered. This fact and the insertion present in segment II indicate the relationship of this group of DNA polymerases from terminal protein containing genomes. On the other hand, the relative evolutionary distances inferred from the DNA polymerase sequences of segment I (Table II) are in good agreement with the qualitative comparisons in segments II and III (see Figure 4), and with the data shown in Table I, when comparing the size of the closely related DNA polymerases.

The results presented indicate that the DNA polymerases compared in this paper are evolutionarily related. Different evolutionary pathways may have been responsible for the relationships detected, such as common ancestry, convergent evolution or transduction of host genes. Both the second and third possibilities seem unlikely since they do not explain the colinearities along the consensus regions of the different DNA polymerases considered. Therefore we favour the hypothesis that the relationships observed are all based on common ancestry, i.e. divergent evolution.

At present there is no evidence that any of these conserved amino acid regions corresponds to functional parts of the DNA polymerases. However, conservation among the distantly related DNA polymerases described here suggests that these regions are parts of functional domains of the enzyme. Recently the C-terminal domain of the HSV DNA polymerase has been proposed to fold, forming the catalytic site of the enzyme that would contain the dNTP, PP and aphidicolin binding sites (Gibbs *et al.*, 1985). The existence of highly conserved domains in this C-terminal region and the similarities in its spatial arrangement suggest a similar three-dimensional structure forming the catalytic domain of the related DNA polymerases described in this paper. It seems reasonable to correlate the similarities in drug sensitivity of the φ29 and T4 enzymes with their structural similarities to the viral α-like DNA polymerases. In particular the high sensitivities to aphidicolin and nucleotide analogues of the T4 DNA polymerase, very similar to those described for viral α-like DNA polymerases, correlate well with the structural homology detected with the polymerases of vaccinia and Herpes viruses. All these analogies provide strong evidence that T4 and animal viruses are somehow evolutionarily related. Interestingly it has recently been reported that the thymidylate synthase gene of phage T4 is distinguished among prokaryotic genes by its expression through an RNA-splicing mechanism analogous to the eukaryotic group I splicing pathway (Ehrenman *et al.*, 1986). In the case of φ29 the fact that the DNA polymerase activity is inhibited by aphidicolin (Blanco and Salas, 1986) and PAA (this paper), whereas the initiation reaction is insensitive, suggests that in addition to the polymerization domain, probably similar in the DNA polymerases compared in this paper, there is a specific initiation domain in the φ29 DNA polymerase.

The structural and functional homologies revealed in these studies suggest that the DNA polymerase may be one of the most phylogenetically stable genes during evolution. As suggested by Ahlquist *et al.* (1985), the adaptation to new hosts or exposure to animal immune systems would be expected to exert more selective pressure on virus structural proteins than on replicases. In this sense the fact that the highly conserved regions described here, although contained in the T4 DNA polymerase, are not present in the sequence of the phage T7 DNA polymerase, suggests the need for caution in thinking about many evolutionary relationships established only by morphological or even host-specificity criteria.

Materials and methods

Replication and initiation assays with phage φ29 DNA–protein p3 as template

The incubation mixture contained, in 25 μl, 50 mM Tris–HCl, pH 7.5, 10 mM $MgCl_2$, 20 mM $(NH_4)_2SO_4$, 1 mM dithiothreitol, 1 mM spermidine, 5% glycerol, dGTP, dCTP, dTTP and [α-^{32}P]dATP as indicated, 0.5 μg of φ29 DNA–protein p3 (Peñalva and Salas, 1982), 20 ng of purified protein p3 (Prieto *et al.*, 1984) and 80 ng of purified protein p2 (Blanco and Salas, 1984). When indicated, PAA (ICN) was added. After incubation at 30°C, the reaction was stopped by adding 10 mM EDTA–0.1% SDS and heating for 10 min at 68°C; the samples were filtered through Sephadex G-50 spun columns (Maniatis *et al.*, 1982) in the presence of 0.1% SDS, and the excluded volume was counted by Cerenkov radiation. When indicated the DNA labeled as described above was denatured by treatment with 0.5 M NaOH and subjected to electrophoresis in alkaline 0.7% agarose gels as described (McDonell *et al.*, 1977) alongside DNA length markers (1 kb ladder from Bethesda Research Labs). After electrophoresis, the length markers were detected with ethidium bromide, and then the gels were dried and autoradiographed with intensifying screens at −70°C. For the initiation assay, the samples were further treated with micrococcal nuclease (Worthington) and subjected to SDS–polyacrylamide gel electrophoresis as described (Peñalva and Salas, 1982). Quantitation was done by excising from the gel the radioactive band corresponding to the p3–dAMP complex and counting the Cerenkov radiation.

DNA polymerase assay with activated DNA as template

The incubation mixture, in 25 μl, was as described above except that 6 μg of activated calf thymus DNA (Cooper Biomedical) was used as template, 5 μM each dGTP, dCTP, dTTP and [α-^{32}P]dATP (1 μCi) were used, and protein p3 was not added. When indicated 0.15 units of the Klenow fragment of *E.coli* DNA polymerase I (Boehringer Mannheim) or 0.08 units of the T4 DNA polymerase (New England BioLabs) were used instead of the φ29 DNA polymerase. Different drugs were added as indicated. After incubation at 30°C, the reaction was stopped, and the samples were filtered through Sephadex G-50 spun columns as described above and the excluded volume was counted by Cerenkov radiation.

Exonuclease assay

The incubation mixture contained, in 25 μl, 50 mM Tris–HCl, pH 7.5, 10 mM $MgCl_2$, 5 mM 2-mercaptoethanol, 9 ng of purified protein p2, and heat-denatured [3′-^{32}P]DNA (10 000 c.p.m.), as described (Blanco and Salas, 1985b). After incubation for 1 min at 37°C, the Cerenkov radiation of the ethanol-soluble material was counted.

Comparison of amino acid sequences of DNA polymerases

The DNA polymerase sequences compared were the following: S1 mitochondrial DNA (Paillard *et al.*, 1985), pGKL1 (Stark *et al.*, 1984), adenovirus type 2 (Alestrom *et al.*, 1982), phage φ29 (Yoshikawa and Ito, 1982), phage pRD1 (Savilahti and Bamford, 1987), phage T4 (E.K.Spicer, unpublished results, GenBank, 1986), vaccinia virus (Earl *et al.*, 1986), HSV 2 (Tsurumi *et al.*, 1987), VZV (Davison and Scott, 1986), EBV (Baer *et al.*, 1984) and HCMV (Kouzarides *et al.*, 1987).

Due to the difficulty of obtaining computer-derived multiple alignments, these were obtained by a series of binary (pairwise) alignments followed by shifting to achieve maximal homology (conservative changes were not considered). Binary alignments were carried out by computer analysis using the programs WORDSEARCH and GAP from the UWGCG (University of Wisconsin Genetics Computer Group), NAR 12 (1) 1984: 387–395.

Acknowledgements

We are grateful to G.Martin for help with the computer analysis and to Drs G.E.Wright and N.C.Brown for the gift of aphidicolin and nucleotide analogues. This investigtion has been aided by research grant 2 R01 GM27242-08 from the National Institutes of Health, by Grant no 3325 from Comisión Asesora para la Investigación Científica y Técnica and by a grant from Fondo de Investigaciones Sanitarias. A.B. and A.Z. were recipients of fellowships from Caja de Ahorros de Madrid and Fondo de Investigaciones Sanitarias respectively.

References

Ahlquist,P., Strauss,E.G., Rice,C.M., Strauss,J.H., Haseloff,J. and Zimmern,D. (1985) *J. Virol.*, **53**, 536–542.
Alestrom,P., Akusjärvi,G., Petterson,M. and Petterson,U. (1982) *J. Biol. Chem.*, **257**, 13492–13498.
Argos,P., Tucker,A.D. and Philipson,L. (1986) *Virology*, **149**, 208–216.
Baer,R., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrel,P.J., Gibson,T.J., Hatfull,G., Hudson,G.S., Satchwell,S.C., Seguin,C., Tuffnell,P.S. and Barrel, B.G. (1984) *Nature*, **310**, 207–211.
Bamford,D.H. and Mindich,L. (1984) *J. Virol.*, **50**, 309–315.
Blanco,L. and Salas,M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5325–5329.
Blanco,L. and Salas,M. (1985a) *Proc. Natl. Acad. Sci. USA*, **82**, 6404–6408.
Blanco,L. and Salas,M. (1985b) *Nucleic Acids Res.*, **13**, 1239–1249.
Blanco,L. and Salas,M. (1986) *Virology*, **153**, 179–187.
Blanco,L., Gutiérrez,J., Lázaro,J.M., Bernad,A. and Salas,M. (1986) *Nucleic Acids Res.*, **14**, 4923–4937.
Carthew,R.W., Chodosh,L.A. and Sharp,P.A. (1985) *Cell*, **43**, 439–448.
Davison,A.J. and Scott,J.E. (1986) *J. Gen. Virol.*, **67**, 1759–1816.
Earl,P.L., Jones,E.V. and Moss,B. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3659–3663.
Ehrenman,K., Pedersen-Lane,J., West,D., Herman,R., Maley,F. and Belfort,M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5875–5879.
Escarmís,C., García,P., Méndez,E., López,R., Salas,M. and García,E. (1985) *Gene*, **36**, 341–348.
Francou,F. (1981) *Mol. Gen. Genet.*, **184**, 440–444.
García,E., Gómez,A., Ronda,C., Escarmís,C. and López,R. (1983) *Virology*, **128**, 92–104.
Gibbs,J.S., Chiou,H.C., Hall,J.D., Mount,D.W., Retondo,M.J., Weller,S.K. and Coen,D.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7969–7973.
Gunge,N., Tamaru,A., Ozawa,F. and Sakaguchi,K. (1981) *J. Bacteriol.*, **145**, 382–390.
Helgstrand,E., Eriksson,B., Johansson,N.G., Lanerö,B., Larsson,A., Misiorny, A., Norên,J.O., Sjöberg,B., Stenberg,K., Stening,G., Stridh,S., Öberg,B., Alenius,S. and Philipson,L. (1978) *Science*, **201**, 819–821.
Hirochika,H. and Sakaguchi,K. (1982) *Plasmid*, **7**, 59–65.
Huberman,J.A. (1981) *Cell*, **23**, 647–648.
Kallin,B., Sternan,L., Saemundssen,A.K., Luka,J., Jornvall,H., Eriksson,B., Tao,P.Z., Nilsson,M.T. and Klein,G. (1985) *J. Virol.*, **54**, 561–568.
Kemble,R.J. and Thompson,R.D. (1982) *Nucleic Acids Res.*, **10**, 8181–8190.
Khan,N.N., Wright,G.E., Dudycz,L.W. and Brown,N.C. (1984) *Nucleic Acids Res.*, **12**, 3695–3706.
Khan,N.N., Wright,G.E., Dudycz,L.W. and Brown,N.C. (1985) *Nucleic Acids Res.*, **13**, 6331–6342.
Kikuchi,Y., Hirai,K. and Hishinuma,F. (1984) *Nucleic Acids Res.*, **12**, 5685–5692.
Kikuchi,Y., Hirai,K., Gunge,N. and Hishinuma,F. (1985) *EMBO J.*, **4**, 1881–1886.
Kitada,K. and Hishinuma,F. (1987) *Mol. Gen. Genet.*, **206**, 377–381.
Knopf,C.W. (1979) *Eur. J. Biochem.*, **98**, 231–244.
Knopf,C.W. (1987) *J. Gen. Virol.*, **68**, 1429–1433.
Kornberg,A. (1980) *DNA Replication*. W.F.Freeman, San Francisco.
Kouzarides,T., Bankier,A.T., Satchwell,S.C., Weston,K., Tomlinson,P. and Barrel,B.G. (1987) *J. Virol.*, **61**, 125–133.
Larder,B.A., Kemp,S.D. and Darby,G. (1987) *EMBO J.*, **6**, 169–175.
Levings,C.S. III and Pring,D.R. (1979) In Scandalios,J.G. (ed.), *Physiological Genetics*. Academic Press, New York, Vol. 5, pp. 171–192.
Lichy,J.H., Nagata,K., Friefeld,B.R., Enomoto,T., Field,J., Guggenheimer,R.A., Ikeda,J.E., Horwitz,M.S. and Hurwitz,J. (1982) *Cold Spring Harbor Symp. Quant. Biol.*, **47**, 731–750.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
McDonell,M.W., Simon,M.N. and Studier,F.W. (1977) *J. Mol. Biol.*, **110**, 119–146.
Moreno,M.A., Carrascosa,A.L., Ortín,J. and Viñuela,E. (1978) *J. Gen. Virol.*, **93**, 253–258.
Moss,B. and Cooper,N. (1982) *J. Virol.*, **43**, 673–678.
Nagata,K., Guggenheimer,R.A. and Hurwitz,J. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 6177–6181.
Nishiyama,Y., Maeno,K. and Yoshida,S. (1983) *Virology*, **124**, 221–231.
Niwa,O., Sakaguchi,K. and Gunge,N. (1981) *J. Bacteriol.*, **148**, 988–990.
Ollis,D.L., Brick,P., Hamlin,R., Xuong,N.G. and Steitz,T.A. (1985) *Nature*, **313**, 762–766.
Paillard,M., Sederoff,R.R. and Levings,C.S. III (1985) *EMBO J.*, **4**, 1125–1128.
Palmer,J.D., Shields,C.R., Cohen,D.B. and Orton,T.J. (1983) *Nature*, **301**, 725–728.
Peñalva,M.A. and Salas,M. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 5522–5526.
Prieto,I., Lázaro,J.M., García,J.A., Hermoso,J.M. and Salas,M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1639–1643.
Pring,D.R., Levings,C.S. III, Hu,W.W.L. and Timothy,D.H. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 2904–2908.
Salas,M. (1983) *Curr. Top. Microbiol. Immunol.*, **109**, 89–106.
Savilahti,H. and Bamford,D.H. (1987) *Gene*, in press.
Sridhar,P. and Condit,R.C. (1983) *Virology*, **128**, 444–457.
Stark,M.J.R., Mileham,A.J., Romanos,M.A. and Boyd,A. (1984) *Nucleic Acids Res.*, **12**, 6011–6030.
Stillman,B.W., Tamanoi,F. and Mathews,M.B. (1982) *Cell*, **31**, 613–623.
Tsurumi,T., Maeno,K. and Nishiyama,Y. (1987) *Gene*, **52**, 129–137.
Watabe,K., Leusch,M. and Ito,J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5374–5378.
Yoshikawa,H. and Ito,J. (1982) *Gene*, **17**, 323–335.

Received on August 26, 1987

Note added in proof

After our paper had been accepted for publication we learned that E.K.Spicer, J.Rush, C.Fung, L.J.Reha-Krantz, J.D.Karam and W.H.Konigsberg are submitting for publication a paper entitled 'Primary structure of T4 DNA polymerase: evolutionary relatedness of prokaryotic and eukaryotic DNA polymerases'.