Nucleic Acids Research, 1996, Vol. 24, No. 24 **4845–4852**

Deoxy- and dideoxynucleotide discrimination and identification of critical 5′ nuclease domain residues of the DNA polymerase I from *Mycobacterium tuberculosis*

Valerie Mizrahi* and Pieter Huberts

Molecular Biology Unit, South African Institute for Medical Research and Department of Hematology, University of the Witwatersrand Medical School, PO Box 1038, Johannesburg 2000, South Africa

Received October 11, 1996; Revised and Accepted November 5, 1996

ABSTRACT

The DNA polymerase I (PolI) from *Mycobacterium tuberculosis (Mtb)* was overproduced in *Escherichia coli* as an enzymatically active, recombinant protein with or without an N-terminal His-tag. The proteins catalysed both the DNA polymerisation of homo- and heteropolymer template-primers and the 5′–3′ exonucleolytic hydrolysis of gapped and nicked substrates but lacked an associated proofreading activity. In accordance with recent predictions [Tabor,S. and Richardson,C.C. (1995) *Proc. Natl. Acad. Sci. USA*, 92, 6339–6343], both recombinant forms of the *M.tuberculosis* enzyme were unable to discriminate against dideoxynucleotide 5′-triphosphates and were thus efficiently inhibited by these chain-terminating nucleotide analogues during DNA synthesis. This unusual property might be potentially exploitable in terms of novel anti-mycobacterial drug design. A mutational analysis of 5′ nuclease domain residues allowed the roles of nine invariant acidic residues to be evaluated. Acidic side chain neutralisation resulted in a ≥20-fold reduction in activity, with the most profound reduction (≥10^4-fold) being caused by neutralisation of the Asp^{125}, Asp^{148} and Asp^{150} residues.

INTRODUCTION

The DNA polymerase I (PolI) from *Escherichia coli* and its large proteolytic (Klenow) fragment (KF) have been the subject of detailed studies which have provided a prototypical molecular model of DNA-dependent DNA polymerisation (1–5). The X-ray crystal structures of KF and its binary complexes have yielded important insights into the architecture of the primer and nucleotide binding sites (2) and have guided on-going mutational studies aimed at elucidating the roles of individual residues in the substrate binding and catalysis of polymerisation (6). The N-terminal 5′ nuclease domain of PolI has also been well characterised at the level of mechanism and substrate specificity (7,8), although the nature of the interaction between the 5′ nuclease and DNA polymerase domains during nick translation synthesis remains obscure. The eubacterial PolI-associated 5′ nucleases share significant sequence homology with the polymerase-independent 5′ nucleases from bacteriophages T4 (designated as T4 RNase H), T5 and T7. Six conserved subdomains containing 14 invariant residues have been identified within this family of prokaryotic nucleases (9). Two of the invariant arginines were found to be essential for catalysis by *Thermus aquaticus (Taq)* PolI (10) and of the nine invariant acidic residues, the Asp^{13}, which is located in a region homologous to that containing a catalytically important aspartate in bacterial RNase HI (11), was also found to be essential for catalysis by *E.coli* PolI (2). The prokaryotic 5′ nucleases are also related to the Rad2/xeroderma pigmentosum complementation group G (XPG) family of proteins (12) at the level of amino acid sequence (13) and function (14). The amino acids essential for substrate binding and for catalysis by a member of the RAD2 family, the human flap endonuclease 1 (FEN-1), have been found to include five acidic residues that are invariant in both 5′ nuclease families (15).

The recently reported X-ray crystal structures of *Taq* PolI (16), T4 RNase H (17) and T5 5′-exonuclease (18) have yielded important insights into the structure–function relationships within this class of enzymes. The *Taq* domain structure revealed a cleft at the bottom of which are clustered seven of the invariant acidic residues within a sphere of 7 Å radius (16). Six of these residues were implicated in the coordination of one high-affinity and two low-affinity metal ions (Zn^{2+} at site I and Mn^{2+} at sites II and III). The pair at sites I and II were sufficiently close (5 Å) to predict a two-metal-ion hydrolysis mechanism for this enzyme analogous to that postulated for DNA polymerases and reverse transcriptases (RTs; 19–22), polymerase-associated 3′–5′ exonucleases (2, 23–25), RNases H (26, 27) and ribozymes (28). The 2.06 Å X-ray crystal structure of T4 RNase H showed a high degree of structural similarity to the *Taq* 5′ nuclease over the region containing the invariant residues, but revealed only two Mg^{2+} ions coordinated at the bottom of a cleft containing all nine of the invariant acidic residues (17). The location and coordination motif of the first Mg^{2+} ion in this structure (site I^{T4}) corresponded closely to that of the site I Zn^{2+} in the *Taq* domain, whereas the site II^{T4} Mg^{2+} ion was located in proximity to the Asp^{200} carboxyl but considerably further removed from the site I metal ion (7 Å). The T5 5′-exonuclease structure revealed a helical arch

*To whom correspondence should be addressed. Tel: +27 11 489 9370; Fax: +27 11 489 9001; Email: 075val@chiron.wits.ac.za

ideally suited for the threading of single-stranded (ss) DNA, at the base of which is the active site containing two metal ions separated by an even greater distance (8.1 Å; ref. 18). The latter two structures have thus cast doubt on the relevance of a two-metal-ion mechanism for the phage nucleases.

We recently reported the cloning and sequence analysis of the *polA* gene from *Mtb* (29) which is closely related to a homologue occurring at two distinct loci in the *Mycobacterium leprae* genome (30). Although the mycobacterial sequences contained all of the residues essential for the catalysis of DNA synthesis (29), a significant sequence deviation was noted in the O helix region of the polymerase domain of the *Mtb* and *M.leprae* polypeptides inasmuch as the Phe residue found in all other bacterial PolI sequences examined to date [Phe^{762} of *E.coli* PolI (6)] was replaced by a Tyr [Tyr^{737} of *Mtb* (29) and Tyr^{740} of *M.leprae* PolI (30)]. Based on the recent study of Tabor and Richardson (31), the existence of a Tyr at this position in the O helix was predicted to abolish the ribose ring selectivity of the enzyme and render it highly sensitive to inhibition by chain-terminating nucleotide analogues. Since this prediction has obvious implications for the design of novel anti-mycobacterial agents, we were prompted to analyse the *Mtb* enzyme, and in this paper we describe the overproduction of two recombinant forms of this protein for use in nucleotide discrimination studies. We also report the use of these enzymes in a mutational study of selected 5′ nuclease residues aimed at identifying functionally and/or structurally important side chains in this domain.

MATERIALS AND METHODS

Materials

Escherichia coli CJ236 was kindly provided by T. Kunkel and *E.coli* HMS174, BL21(DE3) and BL21(DE3)pLysS, for use in the pET expression system (32), were obtained from Novagen (Madison, WI, USA). Restriction enzymes, dNTPs, ddATP and poly(dA) were from Boehringer-Mannheim and $dT_{12–18}$ was from Pharmacia. Radionucleotides were from New England Nuclear. Oligodeoxyribonucleotides (oligos) were synthesised by Genosys Biotechnologies (The Woodlands, TX, USA). HE19-NB (5′-CGGCTTGGCTGGATCCTCGC), HE19-BX1 (5′-GTCCCGGATACGCCGGCC), HE19-BX2 (5′-AGCTGGATCTGCTAGCCAG), HE19-CB (5′-TCCCGCCCCGGATCCACGCTCGG) and HE19-NCOI (5′-TAGTCACCATGGCTCACAC) were used to create expression cassettes of the *polA* gene encoding *Mtb* PolI (29). D21N (5′-AATTGCCATTCAGCAACATC), A70V (5′-CGTCGAAAGCCACCGCGATGTG), A71V (5′-ACACGTCGAAAACCGCCGCGAT), AA7071VV (5′-CGTCGAAAACCACCGCGATGT), D73N (5′-CGGGACACGTTGAAAGCCGC), Q82E (5′-CGGGTAGCGTTCCAAGCGGAA), E123Q (5′-CGTCGGCCTGGAACCCCGG), D125N (5′-TGAGGTCGTTGGCCTCGAA), D126N (5′-CGATGAGGTTGTCGGCCTC), D148N (5′-CGTCACGATTCCCGGTGAC), D150N (5′-GCAGTGCGTTACGATCCCC), D199N (5′-CGCTGGGGTTGCCGCGCAG), D202N (5′-GCAGGTTATTGCTGGGGTC) and Y737F (5′-AGCCAGCCCGAAGGACATCG) were used for mutagenesis of the *polA* gene. D498N (5′-ATAGTAACAAACTCACAAT), RT6 (5′-GGATTCAGGATTAGAAG) and RT7 (5′-GGATTCAGGATTAGAAGTAAAC) were complementary to M13-$RT^{D443Q/D498N}$, an M13 subclone of a RNase H domain mutant of HIV-1 reverse transcriptase (33; denoted M13-RT herein) and were used to construct the double-stranded (ds) DNA substrates used in the 5′ nuclease assays.

Overproduction of *Mtb* PolI in *E.coli*

The cloning and sequencing of the *Mtb polA* gene has been previously described (29). The 3.05 kb C-terminal *Hin*dIII–*Eco*RI fragment of pPH1 was subcloned in M13mp19 to produce M13-HE19. Site-directed mutagenesis of M13-HE19 was carried out as previously described (33,34) using the method of Kunkel *et al.* (35) and Sanger DNA sequencing (36) was performed using Sequenase 2.0 (US Biochemical). The coding region of the *polA* gene was engineered as a *Bam*HI fragment by introducing new sites at the N- and C-termini of the gene and removing the existing *Bam*HI sites within the gene by mutagenesis of M13-HE19 using HE19-NB, HE19-CB, HE19-BX1 and HE19-BX2 to produce $M13\text{-}HE19^{4M}$. The HE19-NB primer created a new *Bam*HI site for in-frame fusion of the pET15b vector to the Ala^{13} residue of the *Mtb* PolI. This position was chosen on the basis of a multiple PolI amino acid sequence alignment which suggested that the first 12 residues of the *Mtb* enzyme would not be required for enzymatic activity. The HE19-CB primer created a new *Bam*HI site 11 bp downstream of the stop codon of the gene and the HE19-BX1 and HE19BX2 primers removed the two internal *Bam*HI sites with silent codon changes [Arg^{662} (CGG→CGA) and Ala^{388} (GCG→GCT), respectively]. The 2697 bp *Bam*HI fragment from $M13\text{-}HE19^{4M}$ was cloned in pET15b to create $pPH3^{His}$, which encoded a fusion protein with the N-terminal sequence: $MGSS(H)_6SSGLVALPRGSHMLEDP\mathit{AKTPL}..$, where the italics denote the *Mtb* PolI sequence, commencing at Ala^{13}. To control for unpredictable effects of this N-terminal His-tag on the catalytic properties on the enzyme, a second construct encoding a non-tagged version of the protein was created by introducing an *Nco*I site in $HE19^{CB/BX1/BX2}$ using HE19-NCOI and cloning the resulting 2727 bp *Nco*I–*Bam*HI fragment in pET15b to create pPH3. This plasmid encoded a protein, V1M, which differed from the native in only the first amino acid (Met in place of Val). Gene expression in *E.coli* BL21(DE3)pLysS was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) and incubation at either 30 (pPH3) or 37°C ($pPH3^{His}$) for a period of 3 h. SDS–PAGE analysis of the induction products confirmed the high-level overproduction of proteins of the predicted size (100 kDa; Fig. 1). Furthermore, the soluble fractions recovered post-induction were found to contain $\geq 10^3$-fold higher DNA polymerase activity than pre-induction controls.

Mutagenesis of the *Mtb polA* gene

In the case of the 5′ nuclease domain mutants, the mutated region was excised as a 1341 bp *Hin*dIII–*Asp*718 fragment and was re-ligated to the 8.5 kb *Hin*dIII–*Asp*718 fragment from wild-type $M13\text{-}HE19^{4M}$ to re-construct the mutant. For the Y737F mutant, the mutation was excised as a 1700 bp *Asp*718–*Eco*RI fragment and was re-ligated to the 8.14 kb *Asp*718–*Eco*RI fragment from wild type $M13\text{-}HE19^{4M}$. Confirmatory sequence analysis was minimised by including these additional cloning steps.

Protein expression and purification

SDS–PAGE analysis of induced cell extracts suggested that the overall expression levels of the wild-type and mutant proteins directed by pPH3 and by $pPH3^{His}$ were similar. However, the

solubility of the non-tagged wild type and D125N proteins was significantly greater than that observed for their His-tagged counterparts (Fig. 1). This difference suggested that the presence of the His-tag favoured the formation of inclusion bodies at 37°C in contrast to the preferential partitioning of the non-tagged proteins in the soluble fraction at the lower induction temperature. However, even within the family of wild type and mutant His-tagged proteins, the partitioning between soluble and insoluble fractions varied between extremes of 10–15% for A70V/A71V and 50% for the wild type (data not shown). The His-tagged wild type, mutants and a control from induced cells harbouring vector alone were purified as follows. Induced cell pellets from 500 ml cultures, resuspended in 10 ml Novagen binding buffer [Buffer A; 20 mM Tris–HCl (pH 7.9), 0.5 M NaCl, 5 mM imidazole] were lysed by one freeze–thaw cycle. The soluble fraction recovered by centrifugation (30 min at 25 000 *g*) was purified by batch-wise metal chelate affinity chromatography using 5 ml resin. The resin was washed twice each with 10 ml Buffer A and salt-free wash buffer [Buffer B; 60 mM imidazole, 20 mM Tris–HCl (pH 7.9)] before eluting the bound protein with 10 ml salt-free elute buffer [Buffer C; 1 M imidazole, 20 mM Tris–HCl (pH 7.9)]. This procedure resulted in the recovery, in a significantly purified form (~50%), of full-length protein in all cases except D73N, E123Q and D126N, where the eluate was found to consist instead of a wide range of smaller polypeptides. However, the extent of degradation during purification was markedly reduced by including Complete^{TM} protease inhibitor cocktail (Boehringer Mannheim) in all buffers, implying that these three mutants were particularly prone to proteolysis. The inhibitors were therefore included in all buffers for the purification of all of the enzymes. The co-purification of a minor *E.coli* nuclease contaminant with the His-tagged recombinant (37) necessitated further purification by anion exchange chromatography. The affinity column eluate was therefore loaded on a 5 ml Q Sepharose (Fast Flow, Sigma) column which was washed with 20 ml Buffer C containing 0.25 M NaCl and eluted with an 80 ml linear gradient from 0.25 to 0.5 M NaCl. The *Mtb* PolI eluting at 0.3–0.35 M NaCl was 80–95% homogeneous. Peak fractions were concentrated 10-fold by centrifugation over Ultrafree MC filters (Millipore) and were stored at 0.05–0.1 mg/ml in a storage buffer [10 mM Tris–HCl (pH 7.9), 1 mM dithiothreitol (DTT), 0.15 M NaCl, 0.5 M imidazole, 50% glycerol]. The activity loss and protein degradation were negligible after 3 months storage at –20°C for all of the proteins except for D126N which degraded rapidly ($t_{0.5}$ = 5 d).

The non-tagged wild-type and D125N mutant proteins were purified as follows. Induced cell pellets from 2.5 l cultures, resuspended in 100 ml salt-free Buffer A and containing protease inhibitors, were lysed as described above and clarified by ultracentrifugation (30 min at 110 000 *g*). The high-speed supernatant was loaded on a 100 ml phospho-cellulose (Whatman P11) column equilibrated in salt-free Buffer B. The column was washed with 40 ml Buffer B and was eluted with a 100 ml linear gradient from 0 to 0.5 M NaCl in Buffer C. The peak *Mtb* PolI-containing fractions eluting at 0.2–0.3 M NaCl were pooled and further purified to >95% homogeneity on a 5 ml Q Sepharose column eluting with an 80 ml linear gradient from 0.25 to 0.5 M NaCl. Peak fractions were concentrated as described above. In all cases, proteins were diluted in 50 mM Tris–HCl (pH 7.0), 150 mM NaCl, 1 mM DTT and 50% glycerol immediately before use in enzyme assays and their concentrations were determined using the BioRad protein assay (Kit I, bovine gamma globulin standard).

DNA polymerase assays

Assays (11 µl) containing 700 nM 3′-OH [poly(dA)·oligo $(dT)_{12–18}$; A:T nucleotide ratio = 860:1], 33 µM [^{3}H]TTP (2500 c.p.m./pmol), 8 mM $MgCl_2$, 5 mM DTT, 8 mM KCl, 15 mM NaCl, 5% glycerol and 16–33 nM PolI in 50 mM Tris–HCl (pH 8.3) were incubated at 37°C for 1–20 min. Samples were quenched and processed by a DE81 filter binding assay (34). For polymerase assays of *E.coli* PolI and KF, reactions were carried out at 37°C in 50 mM Tris–HCl (pH 7.4) containing 2 mM $MgCl_2$.

dNTP/ddNTP discrimination assay

5′-[^{32}P]-D498N was annealed to ss M13-RT at a molar ratio of 1:3. Polymerisation was followed by the gel electrophoretic separation of products generated using a nucleotide mix optimised for T7 DNA polymerase [ddNTP/dNTP ratio of 0.1 (8 µM ddNTP, 80 µM dNTPs); Sequenase version 2.0 kit (US Biochemical)]. Assays containing 25 mM Tris–HCl (pH 8.0), 5 mM $MgCl_2$, 5 mM DTT, template-primer (33 nM 3′-OH), nucleotide mix and enzyme (30–150 nM) were incubated at 37°C for 15 min. Aliquots (1.5 µl) were quenched with 4 µl sequencing gel sample loading buffer and fractionated by electrophoresis in 6% polyacrylamide, 7 M urea, 1× TBE [90 mM Tris, 64.6 mM boric acid, 2.5 mM EDTA, pH 8.3] sequencing gels which were exposed at –70°C for 18 h.

3′–5′ Exonuclease assays

Assays (5 µl) containing 90 nM 5′-[^{32}P]-D498N in 5 mM $MgCl_2$, 2 mM DTT and 50 mM Tris–HCl (pH 8.0) were initiated by the addition of enzyme and incubation at 37°C for 0–40 min. Aliquots (1.5 µl) were quenched and fractionated in 16% sequencing gels, as described above.

5′ Nuclease assays

5′-[^{32}P]-D498N was 3′-ddA-terminated in a 120 µl reaction containing ss-M13-RT/[5′-^{32}P]-D498N (29 nM 3′-OH, ss DNA:primer = 4.3:1), 100 µM ddATP, 6.7 mM DTT, 33 mM NaCl, 13 mM $MgCl_2$ and 10 U Sequenase version 2.0 in 25 mM Tris–HCl (pH 7.5) incubated at 37°C for 1.3 h. The reaction was stopped by PhOH/$CHCl_3$ extraction and was diluted with 680 µl water. Aliquots (100 µl) were purified by centrifugation over 1 ml Sephadex G25 spin columns. The hybrid, M13-RT/(3′-ddA/5′-[^{32}P])-D498N, was resuspended in 60 µl water (47 nM hybrid; 80% recovery) and partitioned into three equal aliquots. The substrates, Gap-∞, Gap-5 and Gap-0, were prepared by annealing either water (Gap-∞) or a 17-fold molar excess of the appropriate unlabelled upstream primer (RT6 or RT7) to this hybrid by heating to 70°C for 1 min and cooling over 1.5 h to 37°C in a buffer containing 48 mM Tris–HCl (pH 8.0) and 4.8 mM $MgCl_2$. Since equivalent concentrations of the various upstream primers were annealed to equal aliquots of the precursor hybrid, the final concentrations of the DNA substrates were the same in all cases. The gap sizes denote the spacing in nucleotides (nt) between the 3′-terminus of the upstream primer (∞, 5 or 0 nt) and the scissile 5′-end of the downstream 3′-ddA-terminated D498N primer.

Complementation studies

The mutant complementation vector, pMTBPOLA^{V1M}, was constructed with the 3.05 kb *Hin*dIII–*Eco*RI fragment from HE19$^{NCOI/CB/BX1/BX2}$ (harbouring the V1M mutation) being

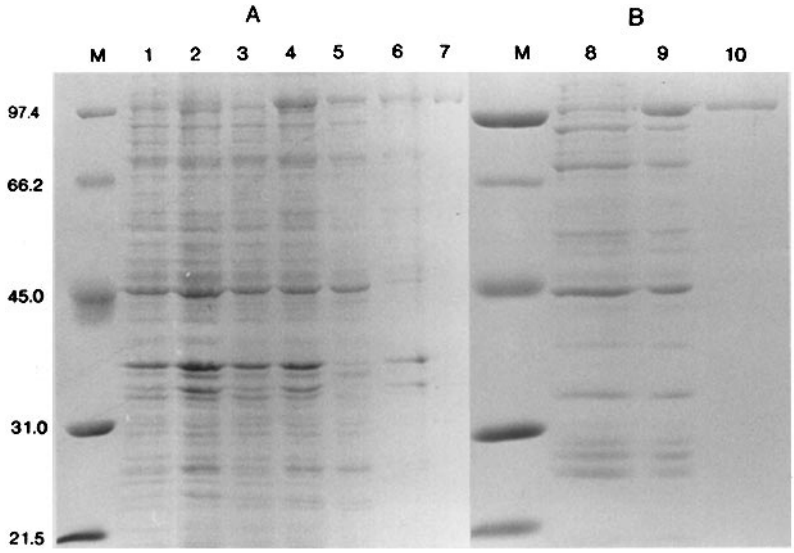

Figure 1. Purification of two recombinant forms of wild type *Mtb* PolI. (**A**) His-tagged protein (lanes 1–7). (**B**) Non-tagged (V1M) protein (lanes 8–10). (A) Lane 1, whole cell extract of uninduced BL21(DE3)pLysS/pET15b; lane 2, whole cell extract of IPTG-induced BL21(DE3)pLysS/ pET15b; lane 3, whole cell extract of uninduced BL21(DE3)pLysS/pPH3^{His}; lane 4, whole cell extract of IPTG-induced BL21(DE3)pLysS/pPH3^{His}; lane 5, soluble fraction of sample shown in lane 4; lane 6, insoluble fraction of sample shown in lane 4; lane 7, purified His-tagged protein. (B) Lane 8, soluble fraction of induced BL21(DE3)pLysS/pET15b; lane 9, soluble fraction of induced BL21(DE3)pLysS/pPH3; lane 10, purified non-tagged protein; M, Molecular weight markers (sizes shown). Samples were fractionated by SDS–PAGE in a 10% gel which was stained with Coomassie Brilliant Blue.

used in place of the corresponding fragment from HE19^{CB} that had been used to create the wild type counterpart, pMTBPOLA (38). The ability of the mutant vector to complement the plasmid replication defect of the *polA755::aph* mutant of *M.smegmatis* (39) was analysed by comparing its transformation frequency to that of the pMTBPOLA (wild type) control (38). All experimental procedures were carried out as previously described (38).

RESULTS

Overproduction and purification of two recombinant forms of *Mtb* PolI in *E.coli*

The plasmid pPH3^{His} directed the overproduction of a His-tagged protein of limited solubility in *E.coli*. However, the protein released in the soluble fraction of the lysate was associated with DNA-dependent DNA polymerase activity and could be purified to homogeneity by a combination of affinity and ion exchange chromatography (Fig. 1A). Since the N-terminal His-tag extension might conceivably affect the enzymatic properties of the protein, a second recombinant form of the enzyme, which differed from that predicted from the gene sequence in only the first amino acid, was also over-expressed. This recombinant form was highly soluble at an induction temperature of 30°C and was purified to >95% homogeneity by ion exchange chromatography (Fig. 1B).

Effect of the V1M mutation on the polymerase activity of *Mtb* PolI expressed in a mycobacterial host

The effect of the V1M mutation on the DNA polymerase activity of *Mtb* PolI, expressed in a heterologous mycobacterial host under the control of its *own* promoter, was analysed by studying the ability of the V1M *polA* gene to complement the plasmid replication defect of a *polA* mutant of *M.smegmatis* insertionally inactivated within its DNA polymerase domain (38). The *Mtb* V1M *polA* gene and flanking sequences were cloned in pOLYG (a shuttle plasmid harbouring the pAL5000 origin of replication and a hygromycin resistance gene) to yield the V1M mutant form of the previously described wild-type complementation vector, pMTBPOLA (38). The vector, pMTBPOLA^{V1M}, transformed *M.smegmatis polA755::aph* to hygromycin resistance with ~30% of the frequency observed for its wild type counterpart (0.3×10^4 c.f.u./μg; ref. 38), confirming that it retained the ability to direct the expression of functional PolI from the *M.tuberculosis polA* promoter *in vivo*. Since the DNA polymerase activity of PolI plays a well established role in the initiation of both continuous- and discontinuous-strand synthesis in the replication of plasmids such as ColE1 (40) and pAL5000 (39), we propose that these data support the validity of using the recombinant V1M protein as a surrogate for studying the biochemical properties of the DNA polymerase activity of the native enzyme. However, we emphasise that this argument is speculative, since it rests on the assumption that the translational initiation site of the *Mtb polA* gene in both *M.tuberculosis* and *M.smegmatis* is indeed at this position [Val in the wild type (29) versus Met in the mutant].

DNA polymerase activity

The DNA polymerase activities of the recombinant *Mtb* PolI enzymes were studied using a homopolymer template-primer. The pH and [Mg^{2+}] optima for the His-tagged protein were found to be 8.4 and 8 mM, respectively (data not shown). The k_{cat} values of the His-tagged and non-tagged forms were 0.4 s^{-1} and 0.6 s^{-1}, respectively, compared to 1.5 s^{-1} for *E.coli* PolI. Since the extent of contamination of the recombinant *Mtb* PolI preparations with inactive protein is unknown, the significance of these relatively low values is unclear. However, the similar kinetics of synthesis by the two recombinant forms suggested that the artificial N-terminal structure of the His-tagged recombinant [27 amino

acid $(His)_6$-bearing peptide fused to Ala^{13}] had a negligible effect on the polymerase activity of the enzyme. This observation is in accordance with the well established fact that the 5′ nuclease and DNA polymerase domains of the *E.coli* and *Taq* PolI enzymes are distinct and separate folding entities (5,41). We therefore concluded that the DNA polymerase properties of both recombinant *Mtb* PolI enzymes would reflect, equally well, those of the native enzyme at a qualitative (but not necessarily quantitative) level.

Nucleotide discrimination during polymerisation

An interesting feature of the PolI enzymes from *Mtb* (29), *M.leprae* (30) and *M.smegmatis* (V. Mizrahi and A.R. De Meyer, unpublished) is the presence of a Tyr residue at a position occupied by a Phe in other bacterial homologues in a region involved in dNTP binding (O helix; Phe^{762} in *E.coli* PolI; ref. 6). We therefore surmised that DNA synthesis catalysed by *Mtb* PolI might be highly susceptible to inhibition by chain-terminators. This hypothesis was tested by monitoring the effect of ddNTPs on the extension of a 5′-end labelled primer annealed to a ss M13 DNA template. To control for potential complications arising from 5′- nuclease-catalysed product degradation, the 5′-nuclease-deficient D199N mutant was analysed in parallel with the wild-type His-tagged enzyme in comparative assays with KF and T7 DNA polymerase using nucleotide mixes optimised for DNA sequencing with the latter enzyme (Fig. 2). In contrast to the high discrimination against ddNMP incorporation demonstrated by KF, the His-tagged *Mtb* PolI was even more susceptible to inhibition of synthesis by ddNMP incorporation than was T7 DNA polymerase. Furthermore, ribose ring discrimination was predictably introduced into the nucleotide binding site of *Mtb* PolI by Tyr^{737}→Phe mutagenesis. Interestingly, the Y737F and Y737F/D199N mutations were poorly tolerated at the level of DNA polymerase activity (6-fold lower than wild type; data not shown), but for the reasons outlined above, the significance of this observation is unclear. The lack of effect of the His-tag on the dideoxynucleotide susceptibility of the DNA polymerase activity of the *Mtb* enzyme was confirmed by analysing the effect of chain terminators on DNA synthesis catalysed by the non-tagged recombinant. A comparison of panels C and H suggested that both forms of the enzyme were indeed equally susceptible to inhibition by ddNTPs.

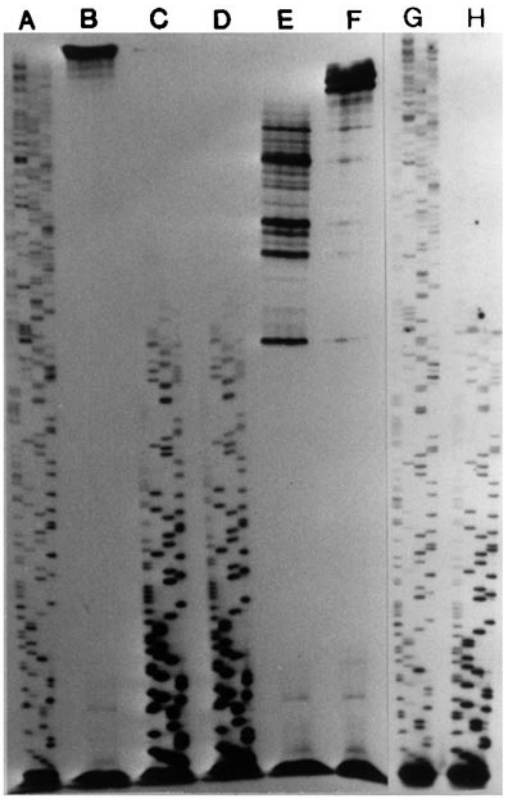

Figure 2. Dideoxy- versus deoxynucleotide discrimination. Assays contained ss-M13-RT/[5′-^{32}P]-D498N in the presence of a T7 DNA polymerase (Sequenase 2.0) nucleotide mix (1:10 ddNTP:dNTP ratio). The four lanes in each panel represent the A, C, G and T tracks generated by T7 DNA polymerase (**A** and **G**), *E.coli* KF (**B**), His-tagged wild type *Mtb* PolI (**C**), His-tagged D199N *Mtb* PolI (**D**), His-tagged Y737F *Mtb* PolI (**E**); His-tagged D199N/Y737F *Mtb* PolI (**F**) and non-tagged wild-type *Mtb* PolI (**H**), respectively.

3′–5′ Exonuclease activity

Exonuclease activity analysis of both recombinant forms of the *Mtb* PolI enzyme using a 5′-end-labelled ss oligo substrate showed negligible 3′–5′ degradation, suggesting that the mycobacterial enzyme lacked an associated proofreading activity (Fig. 3). This property contrasts that of *E.coli* PolI and KF, both of which showed the expected progressive accumulation of 3′-derived hydrolysis products (Fig. 3; ref. 5). The lack of hydrolysis by the *Mtb* enzyme is consistent with the absence of conserved ExoI–III subdomains containing the four critical acidic residues required for 3′–5′ exonuclease activity (42).

5′ Nuclease activity

The activities of both recombinant forms of *Mtb* PolI were then analysed using the three ds DNA substrates described under ‘Materials and Methods’. As shown in Figure 4, the hydrolysis time courses were qualitatively similar for the two enzymes in all cases except for Gap-5, which was hydrolysed more efficiently by the His-tagged enzyme than by its non-tagged counterpart. Although the rate of hydrolysis of Gap-∞ was very low, the positioning of a static upstream primer immediately adjacent to the scissile bond facilitated the formation of a predominantly mononucleotide product to the same extent for both enzymes. Furthermore, in both cases, concomitant extension of an upstream primer towards the scissile 5′-terminus resulted in the formation of an additional minor dinucleotide product (Fig. 4; Gap-5 + dNTPs). These observations were consistent with catalysis by a 5′ nuclease activity anchored via its polymerase domain to an upstream 3′-terminus, with the activation observed in advance of an extending upstream primer being analogous to that reported both for *E.coli* (5,7) and for *Taq* PolI (41). The relatively efficient hydrolysis of Gap-5 by the His-tagged enzyme may have been due to enhanced binding, via the N-terminal His-tag, to a substrate containing a large gap between the upstream anchor site and the scissile 5′-terminus. On the basis of these data, we therefore concluded that the hydrolysis of a *nicked* substrate by the His-tagged recombinant was sufficiently representative of catalysis by a 5′ nuclease to be used as an assay system for probing the effects of mutating selected residues within this domain.

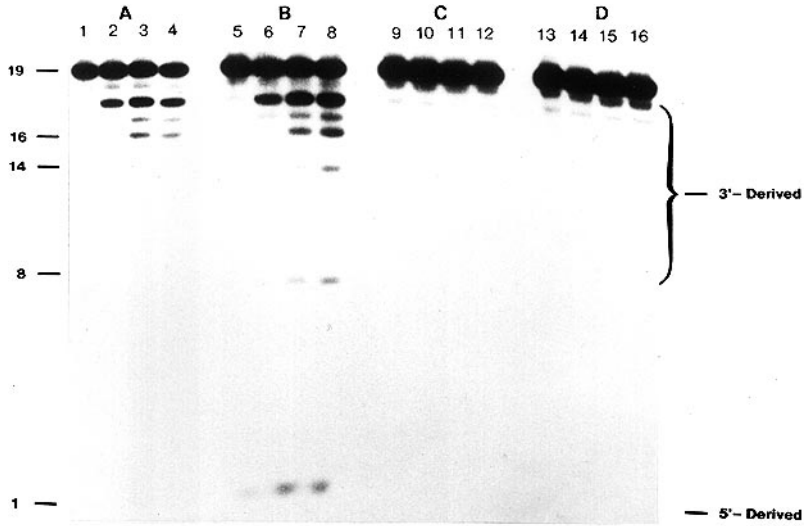

Figure 3. 3′–5′ Exonuclease activity. The activities of *E.coli* KF (**A**), *E.coli* PolI (**B**), His-tagged *Mtb* PolI (**C**) and non-tagged *Mtb* PolI were compared in assays containing [5′-^{32}P]-D498N (90 nM) and enzyme [140 nM in (A–C) and 35 nM in (D)]. Samples were quenched after 0 (lanes 1, 5 and 9), 5 (lanes 2, 6 and 10), 20 (lanes 3, 7 and 11) and 40 min (lanes 4, 8 and 12) and analysed by gel electrophoresis. The sizes and origin of the of the oligos are as indicated.

5′ Nuclease domain mutagenesis

The His-tagged mutant enzymes were purified according to the same procedure as that developed for the corresponding wild type protein (data not shown). As shown in Table 1, the mutations had little or no effect on the specific activity of the DNA polymerase. The 5′ nuclease activities of the mutants were then compared to the wild-type enzyme using the nicked (Gap-0) substrate. Although the A70V, A71V, A70V/ A71V and Q82E mutations had no detectable effect on activity, neutralisation of each of the nine invariant acidic residues reduced the activity by a factor of at least 20-fold with the D125N, D148N and D150N mutations resulting in the most profound reduction in rate ($\geq 10^4$-fold; Table 1). The non-tagged form of the D125N mutant was also purified and analysed and as shown in Table 1, the two forms of this mutant were enzymatically indistinguishable. This observation strongly supported the validity of the remaining mutagenesis data that were obtained using the His-tag expression system alone.

DISCUSSION

The DNA polymerase activity of *Mtb* PolI is highly susceptible to inhibition by chain-terminating nucleotide analogues

The two recombinant proteins described in this study have been used to analyse selected properties of the native *Mtb* PolI enzyme. As described above, the ability of the non-tagged, V1M recombinant to complement the plasmid replication defect of a *polA* mutant of *M.smegmatis* strongly supported the validity of using this protein as a surrogate for studying the DNA polymerase activity of the native enzyme. Since *in vitro* studies suggested that the DNA polymerase activity of the His-tagged recombinant was indistinguishable from that of its non-tagged counterpart, we speculate that this form of the enzyme would also function as a DNA polymerase *in vivo*.

Table 1. Effects of mutation of conserved and invariant 5′ nuclease domain residues on the activity of recombinant *Mtb* PolI

Enzyme[a]	5′ Nuclease (%)[b]	DNA polymerase (%)[c]
His-tagged		
Wild type	100	100
D21N	0.1	90
A70V	100	100
A71V	100	100
A70V/ A71V	100	100
D73N	0.3	90
Q82E	100	100
E123Q	0.5	90
D125N	≤0.01	90
D126N	5	90
D148N	≤0.01	100
D150N	≤0.01	100
D199N	5	100
D202N	0.1	100
Non-tagged		
Wild type	100	100
D125N	≤0.01	100

[a]Protein mutations are abbreviated using the following convention: the number of the mutated amino acid is preceded by the symbol (single-letter acronym) for the wild-type amino acid and is followed by that for the mutant amino acid. Therefore D21N represents the mutant in which the wild-type Asp^{21} residue is replaced by Asn.

[b]Assayed using the nicked substrate in 0–45 min assays containing 25 nM substrate and either 140 nM (His-tagged) or 35 nM (non-tagged) enzyme. Activities, expressed as a percentage of the corresponding wild-type recombinant, were estimated by quantitating the electrophoretically separated mononucleotide product by scanning densitometry of autoradiographs.

[c]Assayed using a poly(dA)·oligo(dT) template-primer and expressed as a percentage of the corresponding wild-type recombinant.

These recombinants were used to confirm the prediction the native enzyme would be unable to discriminate against ddNTP analogues during DNA synthesis (31). The *Mtb* enzyme is therefore

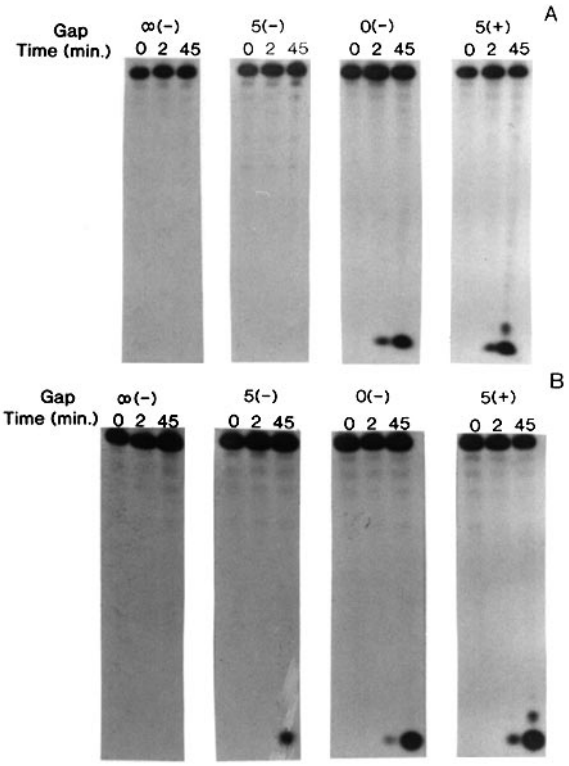

Figure 4. 5′ Nuclease activity. (**A**) Non-tagged wild type *Mtb* PolI. (**B**) His-tagged wild-type *Mtb* PolI. 5′ Nuclease activities were analysed using the three ds DNA substrates described under 'Materials and Methods'. The size of the gap (nucleotides) is shown in the row denoted 'Gap'. Assays containing 25 nM substrate and either 140 nM His-tagged or 35 nM non-tagged enzyme in the presence [denoted as (+) next to the gap size] or absence of dNTPs [denoted as (–) next to the gap size] were quenched at the indicated times (min) and were analysed by gel electrophoresis. As described under 'Materials and Methods', steps were taken to ensure that the ds DNA substrates were produced from ss-M13-RT/(3′-ddA-D498N) in equivalent yields. For a given enzyme, the use of a fixed substrate concentration eliminated the complicating effect that a variable enzyme:DNA ratio would otherwise have had on the relationship between hydrolysis rate and gap size.

the first bacterial PolI described to date that possesses this unusual property. This observation is analogous to that of Eriksson *et al.* (43) which confirmed that the yeast mitochondrial DNA polymerase utilised the chain-terminating anti-HIV agents, ddTTP, 3′-fluoro-TTP and ddCTP almost as efficiently as natural dNTPs. This finding is in turn, consistent with the fact that the mitochondrial polymerase, which is a member of the type I DNA polymerase family, also bears a Tyr residue at the ribose ring discrimination position (6,31). The high degree of primary structural conservation within bacterial PolI enzymes suggests that they may have evolved from a common progenitor which contained a Tyr at the discrimination position. Exposure of rapidly-growing bacteria to naturally occurring chain terminators might have provided the selective pressure necessary for the organisms to evolve a way of discerning the ribose ring structure of the incoming nucleotide. Since the indiscriminate incorporation of chain terminators has potentially catastrophic consequences for bacterial survival, it is reasonable to speculate that the retention of a Tyr residue at this position in the PolI enzymes from both slow- and fast-growing mycobacteria is more likely due to the formidable barrier to entry of naturally occurring chain terminators provided by the unusual mycobacterial cell wall (44) rather than to the intrinsically slow growth rate of these organisms, as has been suggested [ref. 31; the doubling times of *M.smegmatis*, *Mtb* and *M.leprae* are 2–3 h, 18–24 h and 13 d, respectively (45)].

The unusual DNA polymerase enzymology of mycobacterial PolI has obvious implications for anti-mycobacterial drug design. PolI is a repair enzyme required during the growth phase of an organism, since the repair of damage-induced lesions must occur between successive rounds of chromosomal DNA replication. In light of the potential lethality of persistent, unrepaired ss breaks, it is possible that this enzyme is also expressed during the dormant phase of *Mtb*. If this hypothesis is correct, then PolI may be a useful target for chemotherapeutic intervention against dormant tubercle bacilli that are not susceptible to conventional mycobactericidal drugs since these all act specifically against dividing cells. The wide variety of chain-terminating nucleoside analogues developed for anti-viral chemotherapy may therefore include compounds with potential anti-mycobacterial activity. However, the efficacy of such molecules *in vivo* may well be limited by the extent of drug uptake and subsequent metabolism to the triphosphate form.

Structure–function relationships in the 5′ nuclease domain

The 5′ nuclease domain of *Mtb* PolI provided a useful system for analysing structure–function relationships in 5′ nucleases in general. As described above, the efficient 5′–3′ hydrolysis of a nicked substrate catalysed by the His-tagged *Mtb* enzyme provided an appropriate assay for this study. The fact that neutralisation of invariant 5′ nuclease domain acidic residues had a negligible effect on the DNA polymerase activity of *Mtb* PolI argued strongly in favour of independently folded polymerase and 5′ nuclease domains, as has been found for the *E.coli* (5) and *Taq* PolI counterparts (41). However, this same property negated the value of using the DNA polymerase activity as a surrogate marker for mutagenesis-induced global folding defects and hence limited the conclusions that could be drawn from this study to the following:

(i) The effects of neutralisation of the Asp^{21}, Asp^{73}, Asp^{125} and Asp^{148} residues were consistent with their involvement in the coordination of a catalytic site I metal ion, on the following grounds: (a) the requirement of the Asp^{21} residue for catalysis is in agreement with the functional importance of *E.coli* PolI Asp^{13} (2) and FEN-1 Asp^{34} (15), as established by mutagenesis. The *Taq* PolI Asp^{18} is within coordinating distance of the site I Zn^{2+} ion (16) and the T4 RNase H Asp^{19} is bridged to the site I $^{T4}Mg^{2+}$ ion through a water molecule (17). Furthermore, this invariant acidic residue is located in a sequence context related to that found in other nucleases (2), suggesting a functional homology between this residue and the *E.coli* RNase HI Asp^{10} (11) and HIV-1 RNase H Asp^{443} (33, 34) which participate in coordinating the catalytic metal ion in those enzymes. (b) The role of the T4 RNase H Asp^{73} residue in site I metal ion coordination has been established (17) and since the FEN-1 Asp^{86} is also implicated in catalysis (15), the severely reduced activity of the D73N mutant is thus consistent with weakened or loss of site I metal ion binding. (c) The *Taq* PolI and T4 RNase H counterparts of Asp^{125} and Asp^{148} are in close proximity to the site I metal ion, suggesting that these carboxylates participate as ligands for its coordination. However, for reasons that are unclear, the functional homology in the case of the Asp^{148} does not extend to the FEN-1 protein since its counterpart (Asp^{179}) could be neutralised without effect (15).

(ii) The high degree of functional conservation of carboxylates involved in site I metal ion coordination in T4 RNase H and *Mtb* PolI did not extend to the putative site II metal ion ligands, Asp^{150} and Asp^{202}, since the T4 RNase H D157N and D200N mutants retained detectable and wild type activity respectively (17), whereas the *Mtb* PolI D155N and D202N mutants and the corresponding FEN-1 D181A and D233A mutants (15) all had sharply reduced activity. The reasons underlying these differences are presently unclear. However, by analogy with the proposed role of the FEN-1 Glu^{158} and Asp^{233} residues (15), we speculate that the Glu^{123}, Asp^{150} and Asp^{202} residues of *Mtb* PolI are involved in the positioning of a second metal ion which may play an indirect role in substrate binding.

(iii) The extensive degradation of certain mutants during purification in the absence of protease inhibitors suggested that conformational perturbation may have been caused by neutralisation of the Asp^{126}, Asp^{73} and Glu^{123} residues, resulting in increased susceptibility of the corresponding polypeptides to degradation by *E.coli* proteases. By analogy with T4 RNase H, the Asp^{126} residue might participate in a H-bonding interaction with the invariant Arg^{80} residue, and since the T4 RNase H Asp^{71} and Glu^{130} are involved in the positioning of water molecules around the Mg^{2+} ions, neutralisation of Asp^{73} and Glu^{123} may have resulted in structural destabilisation by weakening the metal ion binding and exposing uncoordinated carboxylates to intramolecular electrostatic repulsion.

(iv) The lack of effect of mutations of the conserved Ala^{70}, Ala^{71} and Gln^{82} *Mtb* PolI residues confirmed that they are neither functionally nor structurally important. 'Subdomain B' should therefore be re-defined as a single residue corresponding to Asp^{63} of *E.coli* PolI. Moreover, since the conserved acidic residue in subdomain C [Glu^{72} of *E.coli* PolI (46)] can be neutralised without loss of function, a carboxylate is obviously not required in this subdomain.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support of the Glaxo International Tuberculosis Research Initiative, the South African Medical Research Council, the South African Institute for Medical Research and the Foundation for Research Development. We thank B. Gordhan for performing the complementation experiments. We are deeply indebted to T. Mueser, C. Hyde, N. Nossal and M. Bhagwat for providing us with data prior to publication and for helpful structural insights. We also thank C. Joyce and K. Duncan for advice and encouragement and anonymous reviewers for constructive comments.

REFERENCES

1 Joyce,C.M and Steitz,T.A. (1987) *Trends Biochem. Sci.*, **12**, 288–292.
2 Joyce,C.M. and Steitz,T.A. (1994) *Annu. Rev. Biochem.*, **63**, 777–822.
3 Kuchta,R.D., Mizrahi,V., Benkovic,P.A., Johnson,K.A. and Benkovic,S.J. (1987) *Biochemistry*, **26**, 8410–8418.
4 Dahlberg,M.E. and Benkovic,S.J. (1991) *Biochemistry*, **30**, 4835–4843.
5 Kornberg,A. and Baker,T.A. (1992) *DNA Replication* 2nd. Edn.W.H. Freeman & Co., New York, pp. 108–164.
6 Astatke,M., Grindley,N.D.F. and Joyce,C.M. (1995) *J. Biol. Chem.*, **270**, 1945–1954.
7 Lundquist,R.C. and Olivera,B.M. (1982) *Cell*, **31**, 53–60.
8 Lyamichev,V., Brow,M.A.D. and Dahlberg,J.E. (1993) *Science*, **260**, 778–783.
9 Gutman,P.D. and Minton,K.W. (1993) *Nucleic Acids Res.*, **21**, 4406–4407.
10 Merkens,L.S., Bryan,S.K. and Moses,R.E. (1995) *Biochim. Biophys. Acta*, **1264**, 243–248.
11 Kanaya,S., Kohara,A., Miura,Y., Sekiguchi,A., Iwai,S., Inoue,H., Ohtsuka,E. and Ikehara,M. (1990) *J. Biol. Chem.*, **265**, 4615–4621.
12 Scherly,D., Nouspikel,T., Corlet,J., Ucla,C., Bairoch,A. and Clarkson,S.G. (1993) *Nature*, **363**, 182–185.
13 Robins,P., Pappin,D.J.C., Wood,R.D. and Lindahl,T. (1994) *J. Biol. Chem.*, **269**, 28535–28538.
14 Murante,R.S., Rust,L. and Bambara,R.A. (1995) *J. Biol. Chem.*, **270**, 30377–30383.
15 Shen,B., Nolan,J.P., Sklar,L.A. and Park,M.S. (1996) *J. Biol. Chem.*, **271**, 9173– 9176.
16 Kim,Y., Eom,S.H., Wang,J., Lee,D.-S., Suh,S.W. and Steitz,T.A. (1995) *Nature*, **376**, 612–616.
17 Mueser,T.C., Nossal,N.G. and Hyde,C.C. (1996) *Cell*, **85**, 1101–1112.
18 Ceska,T.A., Sayers,J.R., Stier,G. and Suck,D. (1996) *Nature*, **382**, 90–93.
19 Kohlstaedt,L.A., Wang. J., Friedman,J.M., Rice,P.A. and Steitz,T.A. (1992) *Science*, **256**, 1783–1790.
20 Davies,J.F., II, Almassy,R.J., Hostomska,Z., Ferre,R.A. and Hostomsky,Z. (1994) *Cell*, **76**, 1123–1133.
21 Pelletier,H., Sawaya,M.R., Kumar,A., Wilson,S.H. and Kraut,J. (1994) *Science*, **264**, 1891–1903.
22 Sawaya,M.R., Pelletier,H., Kumar,A., Wilson,S.H. and Kraut,J. (1994) *Science*, **264**, 1930–1935.
23 Freemont,P.S., Friedman,J.M., Beese,L.S., Sanderson,M.R. and Steitz,T.A. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8924–8928.
24 Beese,L.S. and Steitz,T.A. (1991) *EMBO J.*, **10**, 25–33.
25 Zhu,W. and Ito,J. (1994) *Nucleic Acids Res.*, **22**, 5177–5183.
26 Yang,W., Hendrickson,W.A., Crouch,R.J. and Satow,Y. (1990) *Science*, **249**, 1398–1405.
27 Davies,J.F., II, Hostomska,Z., Hostomsky,Z., Jordan,S.R. and Matthews,D.A. (1991) *Science*, **252**, 88–95.
28 Steitz,T.A. and Steitz,J.A. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 6498–6502.
29 Huberts,P. and Mizrahi,V. (1995) *Gene*, **164**, 133–136.
30 Fsihi,H. and Cole,S.T. (1995) *Mol. Microbiol.*, **16**, 909–919.
31 Tabor,S. and Richardson,C.C. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 6339–6343.
32 Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Duberndorff,J.W. (1990) *Methods Enzymol.*, **185**, 60–89.
33 Mizrahi,V., Brooksbank,R.L. and Nkabinde,N.C. (1994) *J. Biol. Chem.*, **269**, 19245–19249.
34 Mizrahi,V., Usdin,M.T., Harington,A. and Dudding,L.R. (1990) *Nucleic Acids Res.*, **18**, 5359–5363.
35 Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) *Methods Enzymol.*, **154**, 367–382.
36 Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
37 Hengen,P.N. (1995) *Trends Biochem. Sci.*, **20**, 285–286.
38 Gordhan,B.G. and Mizrahi,V. (1996) *Gene*, in press.
39 Gordhan,B.G., Andersen,S.J., De Meyer,A.R. and Mizrahi,V. (1996) *Gene*, in press.
40 Kornberg,A. and Baker,T.A. (1992) *DNA Replication* 2nd Edn.W.H. Freeman & Co., New York, pp. 641–645.
41 Lawyer,F.C., Stoffel,S., Saiki,R.K., Chang,S.-Y., Landre,P.A., Abramson,R.D. and Gelfand,D.H. (1993) *PCR Methods Appl.*, **2**, 275–287.
42 Blanco,L., Bernad,A. and Salas,M. (1992) *Gene*, **112**, 139–144.
43 Eriksson,S., Xu,B. and Clayton,D.A. (1995) *J. Biol. Chem.*, **270**, 18929–18934.
44 Brennan,P.J. and Nikaido,H. (1995) *Annu. Rev. Biochem.*, **64**, 29–63.
45 Wheeler,P.R. and Ratledge,C. (1994) In Bloom,B.R. (ed.), *Tuberculosis: Pathogenesis, Protection and Control.* ASM Press, Washington, D.C., pp. 353–385.
46 Joyce,C.M., Kelly,W.S. and Grindley,N.D.F. (1983) *J. Biol. Chem.*, **257**, 1958–1964.